ION PAGE Dist: A

Form Approved
OMB No 0704-0188

(1)

...hour per response, including the time for reviewing instructions, searching existing data sources, ...e collection of information. Send comments regarding this burden estimate or any other aspect of this ...Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ...anagement and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

E

3. REPORT TYPE AND DATES COVERED
FINAL 01 Jun 91 TO 31 May 94

TITLE AND SUBTITLE
DEVELOPMENT OF TECHNIQUES FOR THE EVALUATION OF TOXICANT IMPACTS TO MULTISPECIES SYSTEMS

5. FUNDING NUMBERS
AFOSR-91-0291
61103D
3484/RS

AUTHOR(S)
Dr Wayne G. Landis

PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute of Environmental Toxicology & Chemistry
Huxley College of Environmental Studies
Western Washington University
Bellingham, WA 98225

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR- 94 0524

SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr Walter Kozumbo
AFOSR/NL
110 DUNCAN AVE SUITE B115
BOLLING AFB DC 20332-0001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

1. SUPPLEMENTARY NOTES

2a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release;
distribution unlimited.

A

12b. DISTRIBUTION CODE

3. ABSTRACT (Maximum 200 words)

In this research program, new methods of data analysis were applied to the analysis of multispecies toxicity tests using three complex toxicants. The water soluble fraction of the turbine fuels Jet-A, jP-4 and jP-8 have been examined as stressors for two microcosm protocols, the standardized aquatic microcosm (SAM) and the mixed flask culture (MFC). The SAM is a 3 L system inoculated with standard cultures of algae, zooplankton, bacteria, and protozoa. In contrast, the MFC is 1 and is inoculated with a complex mixture of organisms derived from a natural source. Analysis of the organism counts and physical data were conducted using conventional and newly derived multivariate nonmetric clustering methods and computer visualization techniques. Several fundamental discoveries regarding the impacts of toxicants on ecological systems were made. The first is that recovery of an ecosystem in the sense that it returns to the original or reference state is not a property of these systems. In fact, it is unlikely that recovery is a property of other large ecological systems. In our experiments the various treatment groups incorporated the information as to toxicant concentration that was expressed after periods of so-called recovery. The differentiation of the treatment groups occurred even after the

4. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
U	U	U	U

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF COLOR PAGES WHICH DO NOT REPRODUCE LEGIBLY ON BLACK AND WHITE MICROFICHE.

CONTINUE OF 298 ABSTRACT (AFOSR-91-0291)

elimination of the toxicant from the test system. Another fundamental discovery is that multispecies toxicity tests are not repeatable, although within one experiment the replicates of a treatment group are replicable. In other words, initial conditions are important. The outcome of this research may lead to a new viewpoint in describing the impacts of toxicants on complex ecological systems. This viewpoint is described as the Community Conditioning Hypothesis.

Development Of Pattern Recognition Techniques for the Evaluation of Toxicant Impacts to Multispecies Systems

Wayne G. Landis
Robin A. Matthews
Institute of Environmental Toxicology and Chemistry
Huxley College of Environmental Studies
Western Washington University
Bellingham, WA 98225
206/650-6136

Geoffrey B. Matthews
Computer Science Department
Western Washington University
Bellingham, WA 98225
206/650-3797

Final Technical Report
1 June 1991 - 31 May 1994
USAFOSR
Grant No. AFOSR-91-0291 DEF

28 June 1994

UNCLASSIFIED

Development Of Pattern Recognition Techniques for the Evaluation of Toxicant Impacts to Multispecies Systems

USAFOSR
Grant No. AFOSR-91-0291 DEF

Wayne G. Landis
Robin A. Matthews

Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA 98225, 206/650-6136

Geoffrey B. Matthews

Computer Science Department, Western Washington University, Bellingham, WA 98225, 206/650-3797

28 June 1994
Final Technical Report for Period 1 June 1991 - 31 May 1994

UNCLASSIFIED

Abstract -- In this research program, new methods of data analysis were applied to the analysis of multispecies toxicity tests using three complex toxicants. The water soluble fraction of the turbine fuels Jet-A, JP-4 and JP-8 have been examined as stressors for two microcosm protocols, the standardized aquatic microcosm (SAM) and the mixed flask culture (MFC). The SAM is a 3 L system inoculated with standard cultures of algae, zooplankton, bacteria, and protozoa. In contrast, the MFC is 1 L and is inoculated with a complex mixture of organisms derived from a natural source. Analysis of the organism counts and physical data were conducted using conventional and newly derived multivariate nonmetric clustering methods and computer visualization techniques. Several fundamental discoveries regarding the impacts of toxicants on ecological systems were made. The first is that recovery of an ecosystem in the sense that it returns to the original or reference state is not a property of these systems. In fact, it is unlikely that recovery is a property of other larger ecological systems. In our experiments the various treatment groups incorporated the information as to toxicant concentration that was expressed after periods of so-called recovery. The differentiation of the treatment groups occurred even after the elimination of the toxicant from the test system. Another fundamental discovery is that multispecies toxicity tests are not repeatable, although within one experiment the replicates of a treatment group are replicable. In other words, initial conditions are important. The outcome of this research may lead to a new viewpoint in describing the impacts of toxicants on complex ecological systems. This viewpoint is described as the Community Conditioning Hypothesis.

Key Words: Multispecies toxicity test, Standardized Aquatic Microcosm, Mixed Flask Culture, Non-metric clustering and association analysis, non-equilibrium dynamics

Program Summary, 1991-1994

A common assumption in environmental toxicology is that after the initial impact, ecosystems recover to resemble the control state. This assumption may be based more on our inability to observe an ecosystem with sufficient resolution to detect differences, than reality. Recent findings of complex and perhaps chaotic dynamics in two relatively simple types of microcosms demonstrate that complex dynamics and non-equilibrium systems are the rule rather than the exception.

In the Standardized Aquatic Microcosm and the Mixed Flask Culture (MFC) microcosms, multivariate analysis and clustering methods derived from artificial intelligence research was able to differentiate oscillations that separate the treatments from the reference group, followed by what would normally appear as recovery, followed by another separation into treatment groups as distinct from the reference treatment. The explanation may be that the oscillations are the result of the intrinsic chaotic behavior of population interactions, of which the alteration of detrital quality is but one of many. In fact, preliminary data indicate that material derived from the jet fuel may be released back into the water column due to the decay or organic material. The initial impact of the toxicant re-set the dosed communities into different regions of the n-dimensional space where recovery may be an illusion due to the incidental overlap of the oscillation trajectories occurring along a few axes.

We now use the new visualization technique of space-time worms to see the trajectories of the ecosystems through n-dimensional ecosystem space. The dynamics appear to have little regularity and

resemble chaotic systems in the lack of repeatability and the importance of initial conditions. The dynamics of ecosystems may be more closely related in terms of basic dynamics to such phenomena as turbulence and weather formation. The implications for risk assessment and resource management are being examined.

Program Objectives

The principal objective of this project is to examine the patterns in toxicity data from experiments using two microcosm protocols. We use nonmetric clustering, a multivariate pattern recognition technique developed by Matthews and Hearne (1991), for our primary pattern analyses. NMC has been shown to work well on a variety of ecological data sets (Matthews and Hearne, 1991). The results from the NMC analyses are then compared with those from other standard multivariate techniques to compare the utility of each technique for analyzing aquatic toxicity data.

Specific objectives of the program were:

- Conduct one series of toxicity tests using the SAM and Mixed Flask Culture (MFC) protocols with 3 complex toxicants such as the water soluble fraction of JP-4, shale derived JP-4, and JP-8.
- For at least one of the complex toxicants, conduct a second complete series of toxicity tests (SAM and MFC) to compare similarities between parallel tests.
- Examine the SAM and MFC complex toxicant data using NMC, linear discriminant analysis, correspondence analysis, and metric clustering (k-means using Euclidean and cosine distances).
- Examine existing SAM data from experiments conducted previously for copper sulfate, brass, and graphite using NMC, linear discriminant analysis, correspondence analysis, and metric clustering.
- Describe a protocol that can be used for analyzing multispecies toxicity data. This protocol will incorporate a discussion of the advantages and limitations of the different multivariate analytical tools that were tested during this project.

We have been able to meet each of these objectives and also to develop an important hypothesis that describes the effects of stressors upon ecological systems. The Community Conditioning Hypothesis may be an important key in understanding the ramifications of toxicant impacts at the molecular and ecosystem levels.

Status of the Research

The results from the three years of the research program have been presented at the Annual Meetings of the Society for Environmental Toxicology and Chemistry (SETAC), the 1993 First SETAC World Congress in Lisbon, Portugal, and the yearly Symposium for Environmental Toxicology and Risk Assessment sponsored by Committee E47 of the American Society for Testing and Materials (ASTM). In addition to these presentations, we have also presented our research results during several invited seminars, including the Keynote Address, "Ecosystem Dynamics: Wormspace, Chaos and the Implications for Ecological Risk Assessment", USEPA Regional Risk Assessment Annual Meeting, May 4, 1993, Atlanta, GA.

Since September 1991, we have also prepared and submitted nine manuscripts, three of which have appeared in publication. Copies of these papers are presented in Appendix A.

In the three year program, the specific accomplishments met include:

- Completing SAM experiments using Jet-A, JP-4 and two JP-8 experiments. The second JP-8 SAM was twice the duration of the typical experiment.
- Completing MFC microcosm experiments using the standard protocol for the toxicants Jet-A , JP-4 and JP-8.
- An extensive investigation into the degradation of the WSF materials in the SAM and MFC systems has led to the preliminary conclusion that the biological communities may release these materials into the media during decomposition, redosing the system.
- Completing three sets of MFC experiments modified to explore specific questions as to the design of multispecies toxicity tests.
- Derivation of a novel method to examine ecological dynamics at the community and ecosystem level, the space-time worms.
- Incorporation of nonlinear dynamics and chaos into the interpretation of ecosystem dynamics due to anthropogenic inputs.
- Improvements to the RIFFLE program, providing a graphical user interface so that nonmetric clustering and its association analysis can be accomplished without extensive programming.
- Application of these results to ecological risk assessment, including the conclusion that risk assessments are more akin to weather forecasts, that is forecasts with specified time limits that deal with a chaotic system.
- Derivation of the Community Conditioning Hypothesis, a new means of understanding the changes and dynamics of ecological systems stressed by xenobiotics.
- Examination of databases distinct from our typical research has also proven fruitful. RIFFLE and other multivariate tests were useful in determining biomarker patterns in two sets of data, a sea anemone toxicity test with copper sulfate, and a USEPA field experiment using molecular markers derived from voles exposed to pesticides.
- The technology transfer program has also proven successful with the methods developed as part of this grant which are currently being adopted for the evaluation of effluents from refineries, assessing the long-term impacts of the Exxon Valdez spill, and in the understanding of risks associated with engineered organisms. A short course was presented at SETAC 93, that exposed 35 individuals to multivariate analysis and the use of AI in data visualization. A similar mini course was held in the spring of 1994 for researchers at NOAA's Sandy Point Laboratory. In order to conduct these workshops, manuals that describe the techniques developed by this research have been printed and are included in Appendix B.

The research described above has been presented in a number of peer reviewed publications, Master's theses, technical reports, and course manuals. These writings are attached as Appendices. Below is a summary of our research program from June 1, 1991 to May 31, 1993 with an emphasis on year three of the program.

Nonmetric Clustering, Association Analysis (NMCAA), and Space-time Worms

Unlike the more conventional multivariate statistics, nonmetric clustering is an outgrowth of Artificial Intelligence (AI) and a tradition of conceptual clustering. In this approach, an accurate description of the data is only part of the goal of the statistical analysis technique. Equally important is the intuitive clarity of the

resulting statistics. For example, a linear discriminant function to distinguish between groups might be a complex function of dozens of variables, combined with delicately balanced factors. While the accuracy of the discriminant may be quite good, use of the discriminant for evaluation purposes is limited because humans cannot perceive hyperplanes in highly dimensional space. By contrast, conceptual clustering attempts to distinguish groups using as few variables as possible, and by making simple use of each one. Rather than combining variables in a linear function, for example, conjunctions of elementary "yes-no" questions could be combined: species A greater than 5, species B less than 2, and species C between 10 and 20. Numerous examples throughout the artificial intelligence literature have proven that this type of *conceptual* statistical analysis of the data provides much more useful insight into the patterns in the data, and is often more accurate and robust. Delicate linear discriminants, and other traditional techniques, chronically suffer from overfitting, particularly in highly dimensioned spaces. Conceptual statistical analysis attempts to fit the data, but not at the expense of a simple, intuitive result. Patterns detected by the clustering are then tested against the hypothesized pattern using association analysis. A more detailed description of nonmetric clustering and association analysis has been published (Matthews and Hearne, 1991) and a brief outline of our multivariate methods can be found in Appendix A.

The use of nonmetric clustering in the analysis of ecological datasets has led us to formulate the community conditioning hypothesis. The community conditioning hypothesis states that ecological communities tend to preserve information about every event in their etiology. In our studies of standardized aquatic microcosms (SAMs), for example, we observed distinct community changes in response to stress that would appear and disappear over a two-month period (Landis et al., 1993b; Landis et al., *in press*). Thus, even after the dosed systems had "recovered" to a state indistinguishable from the reference systems, a stress effect reappeared. A purely stochastic system could not exhibit this effect, since information is erased over time and two systems with identical distributions will remain identically distributed. A chaotic system could exhibit this effect, but we do not believe these microcosms are inherently chaotic, since similar systems tend to follow similar evolutions, without the divergences characteristic of nonlinear systems. Instead, we advanced the hypothesis that an unobserved feature of the community carried information about the stressor throughout the history of the system. In the case of the SAMs, we hypothesize detrital conditioning as the mechanism by which information is preserved. We are currently engaged in research testing this specific hypothesis. However, in general, it will be difficult or impossible to observe or predict the mechanisms (genetics, competitive interactions, migration dynamics, community structure, etc.) that will preserve information for an ecological community.

In cases like these, attempting to predict risk for such systems with physical models will be impossible. Useful physical models must be deterministic, stochastic, or chaotic, and our hypothesis rules out each of these. Instead, what is needed to predict risk for such systems will be a tool that analyzes them in a manner more similar to the human expert. In the AI literature, a contrast is made between *similarity-based* systems and *explanation-based* systems (Lebowitz, 1990). A traditional physical model, for example, which will incorporate each relationship in the real world into a relationship between objects in the computer program, is an explanation-based system. It attempts to reconstruct the cause and effect evolution of the real-world system

within the computer system and it will account for the observed data by explaining its causes. Expert systems are also explanation-based systems, but are closely tied to the explanations given by human experts. Similarity-based systems stand in contrast to these systems. Similarity-based systems attempt to analyze the data on its own terms, without preconceptions about explanations. Generally, similarity-based systems attempt to discover abstractions or generalizations that can reduce the complexity of the data. Similarity-based systems excel in discovering patterns and relationships within the data that were unknown to human experts, and have in fact been used with great success to diagnose soybean diseases, discover new classes of stars, and design aircraft subsystems (Michalski and Chilausky, 1980; Cheeseman et al., 1988; Domeshek et al., 1994).

Projections and Space-time Worms

One inherent difficulty of understanding multivariate data such as those from microcosm experiments is the problem of visualization. However, projections of the hyperdimensional data into three dimensional data for visualization may be valuable in describing the relative positions and dynamics of the experimental groups from a laboratory microcosm or field experiment.

Nonmetric clustering can give some help in determining appropriate projections--the variables or parameters that are the most associated with the clustering are obvious candidates for a projection. However, if there are more than two or three of them, we have a (reduced) version of the same problem. As a result, some linear projections, such as Principal Components Analysis (PCA) or Covariance Analysis (COA) might be useful as a further insight into the nature of the patterns in the data. Each of these methods is actually a version of "projection pursuit", in its full generality: seek a projection of the data that maximizes some property of the data. PCA, for example, maximizes covariance or correlation.

Presently, we are working on a version of projection pursuit that maximizes the nonmetric associations we have seen, above. Instead of looking at the scatter plot matrix, projections onto all the original axes, and measuring the association in the quantile quadrats, we are working on an algorithm that will look at the association for quantile quadrats in an arbitrary projection. There is little mathematical theory to guide such a search, so it necessarily has to be heuristic. However, we have some promising early results which show that a good projection can be found reliably in reasonable time. Such a projection would be an adjunct to the standard projections, and reveal different patterns in the data.

A final problem confronting long-term experimentation is the integration of time into the analysis. Observations taken on the same system over a period of time are obviously correlated, so the analyst has the choice of investigating each day individually, and then combining the analyses, or analyzing all of the days together, but taking care that the time-correlations are considered. Time-series analysis is little help, because it is almost exclusively concerned with univariate changes over time - cycles, trends, *etc.* With a multivariate system changing over time, there is no such thing as going "up" or "down", there is only "hither" and "yon". There are a great many directions to go in a 15 or even higher dimensional space.

One way of visualizing this day-to-day change in a two-dimensional projection of the data is with a three-dimensional, interactive computer graphic of the resulting space-time "worm": the cylindrical surface

generated by the two data dimensions and time. We have implemented a graphical tool, and one example of how the data look is shown in Figure 1 with two of the treatments from a jet fuel microcosm projected. Three-dimensional space-time worms can depict two-dimensional dynamics of ecological systems and allow better comparisons than traditional, one-dimensional graphs.

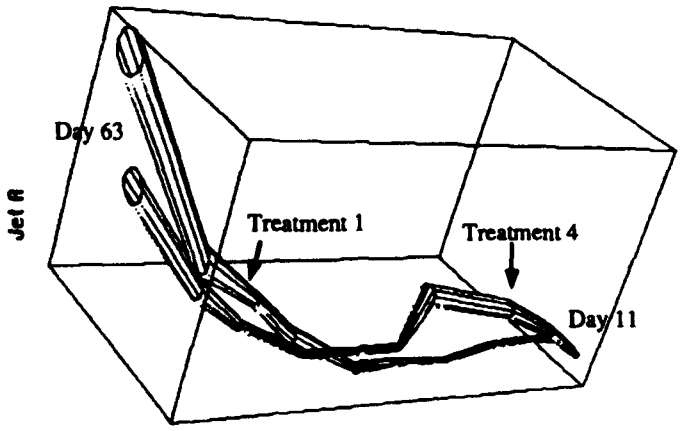


Figure 1. Space-time Worms of the Jet-A Experiment. Although the experiments begin at very similar points, the dynamics of the system are quite different during the course of the experiment. At the midpoint of the experiment the treatments actually converge, and apparently pass through the same area. However, the systems again diverge and are quite distinct at the end of the 63 day long experiment. The axes in this projection have been determined by a PCA analysis of 15 measure biotic variables.

The space-time worms certainly depict the intrinsic dynamics of ecological systems and allows comparisons to be made within ecosystem space. These projections allow us to visualize the relative dynamics between treatments of one experiment and also to compare the various multispecies toxicity tests. Several of the color renditions are presented in Appendix C. Coupled with the tests of significance derived from metric and nonmetric clustering, these projections allow a new and powerful means of describing impacts to ecological systems.

The tools developed as part of this research program have allowed a dramatic increase in our ability to resolve differences in ecological systems. Our similarity-based systems of the analysis of microcosm experiments and the visualization technique of space-time worms has led to the development of the community conditioning hypothesis.

The Community Conditioning Hypothesis

A common assumption in ecological risk assessment is that after the initial stress, ecosystems recover to resemble the control state or reference site. In some instances a new equilibrium state may be established and these dynamics can be described in probabilistic terms (Bartell et al., 1992).

In our series of microcosm experiments using jet fuels as toxicants, analyses as described above were able to differentiate oscillations that separate the treatments from the reference group, followed by what would normally appear as recovery, followed by another separation into treatment groups as distinct from the

reference treatment. Thus, even after the dosed systems had "recovered" to a state indistinguishable from the reference systems, a stress effect reappeared.

The community conditioning hypothesis states that ecological communities tend to preserve information about every event in their etiology. A purely stochastic system could not exhibit this effect, since information is erased over time and two systems with identical distributions will remain identically distributed. A chaotic system could exhibit this effect, but we do not believe these microcosms are inherently chaotic, since similar systems generally tend to follow similar evolutions, without the divergences characteristic of nonlinear systems. Instead, we advance the hypothesis that an unobserved feature of the community carried information about the stressor throughout the history of the system. In the case of the SAMs, we hypothesize detrital conditioning as the mechanism by which this information is preserved. Indeed, degradation rates of the jet fuel materials within the microcosms does seem to be altered upon redosing depending upon prior treatment (Markiewicz, 1994). The information of the etiology of the system can be carried in the structure of the community, the population dynamics of the constituents, and in the structure of the genomes of the populations.

The fact that complex systems have historical components that determine future events has been recognized (Nicolis and Prigogine, 1989). The coevolution of the genetic elements and the response of the resultant community to stressors has been explored by Kauffman and Johnsen (1991) and Kauffman (1993). However, we believe that community conditioning has distinctive properties relative to these constructs. The above theories rely on physical simulation models, models that may predict population dynamics, but have difficulty in generating conceptual shifts in the structure of the ecological system. In a similarity based model, the data are examined to discover abstractions or generalizations that can reduce the complexity of the data. Unknown patterns and relationships are often found and have proven useful in medicine, engineering and astronomy.

The community conditioning hypothesis generates specific and testable hypotheses regarding ecosystems at the community level. Evolutionary events, whether the introduction of a new species, gene or other stressor, are incorporated as the "memory" of the ecosystem. This structure along with the exact nature of the disturbance must be incorporated into the etiology of the outcome, that according to the community conditioning hypothesis, may be widely separated from the initial stressor event. Specific hypotheses generated from Community Conditioning include:

1) The complexity and nonlinear dynamics of a biological community may create long latency periods between observable cause and effects and these effects may be a categorical change in the structure of a community.

2) There are patterns in common to communities of different compositions and physical scales. These patterns are more likely to be those at the system level rather than particular interactions among species.

3) The history of an ecosystem is essential in determining the etiology of an effect due to a toxicant or other stressor. In other words, recovery to an optimal or ground state does not occur.

The accumulation of additional data dealing with changes at the molecular level should enable our research group to better describe community conditioning and contrast it to the traditional recovery and

stability models. The ability to describe clusters in unique ways, to visualize the dynamics of ecological systems using sophisticated projection techniques and then to examine the dynamics for similarities provides a unique opportunity for the examination of these hypotheses.

Comparison of the Standardized Aquatic Microcosm and the Mixed Flask Culture Systems

The experimental designs of the two methods (Table 1) reveal a great deal of similarity. The numbers of groups and the replicates in each group are identical with a total of 24 experimental units available for analysis. The reinoculation of the SAM with algae and other taxa to simulate migration during the course of the experiment is not performed in the MFC. The greatest difference in the designs is the fact that the SAM system is inoculated with set amounts of organisms, minimizing historical inputs before the introduction of the toxicant. In the MFC protocol, a naturally derived inoculum is used. This inoculum is typically a combination of several collections and a three month maturation period occurs before samples are withdrawn for the test procedure. As the experimental units are constructed, a maturation period of 6 weeks is allowed with cross inoculation among the experimental units performed. Cross inoculation stops at the time of toxicant addition. This method allows for a greater number of species, many rare, and also sets each unit with it own historical identity.

In the physical construction of the microcosm units (Table 2) the systems are again similar. Total volume of the SAM is maintained at 3 K while the MFC is 950 mL of media. Not only is there less volume in the MFC, but a calculation of the surface of the container to volume ratio indicates that the MFC has 1.5 times the surface to volume ratio of the SAM method. Organisms and fate processes that are located on the glass surface and sediment are likely to occur at different rates in two systems.

The types of measurements taken as part of the SAM and MFC protocols are similar (Table 3). The biggest difficulty and difference is that in the MFC, with its larger number of species, it is difficult to identify the organisms to species level within a reasonable work load. Because of this, many groupings are combined as in Total Ciliates or Other Bluegreen Algae. The resolution of structure is therefore not as detailed as in the SAM protocol. On the other hand it may be argued that the SAM method has less structure because of its lower number of species.

A list of our data analysis techniques that are used for both methods are listed in Table 4. The comparisons made here concentrate upon the NMCAA tool, but other methods are available. Again, the very different structures of the systems can affect the data analysis. The occurrence of numerous species in the MFC, many of them rare, can make conventional data analysis difficult since rare organisms may be absent in many of the sample collections.

Table 1. Comparison of the experimental designs of the SAM and MFC multispecies toxicity tests. The numbers of groups and replicates are identical in each system.

Experimental Design

Standardized Aquatic Microcosm	Mixed Flask Culture
Number of groups: 4	Number of groups: 4
Number of replicates: 6	Number of replicates: 6
Reinoculation: Once per week add one drop (circa 0.05 mL) to each microcosm from a mix of the ten species = 5 x 102 cells of each alga added per microcosm	Reinoculation: Only reinoculated and cross innoculated during the maturation period.
Addition of test materials: Add material on Day 7	Sampling frequency: 2 times each week
Sampling frequency: 2 times each week	Test duration: 6-8 weeks Allow to mature 6 weeks prior to treatment; track 6 to 8 weeks after exposure. Microcosms are rotated once a week in the environmental chamber during the experiment.
Test duration: 63 days	

Table 2. Comparisons of the physical and chemical structure of the SAM and MFC multispecies toxicity tests. The media are identical except for the addition of $NaHCO_3$ in the MFC protocol. Due to the reduced volume of the MFC and its container, the MFC has 1.5 times the surface to volume ratio of the SAM experimental unit.

Size, Medium and Sediment

Standardized Aquatic Microcosm	Mixed Flask Culture
One-gallon (3.8 L) glass jars are recommended; soft glass is satisfactory if new containers are used; measurements should be 16.0 cm wide at the shoulder, 25 cm tall with 10.6 cm openings.	1 L beakers covered with a large petri dish
Microcosm medium: 3 L T82MV	Microcosm medium: 900 mL of T82MV supplemented with 15 µg $NaHCO_3$ as an additional carbon source, into which 50 mL of inoculum was introduced
Sediment: Composed of silica sand (200 g), ground, crude chitin (0.5g), and cellulose powder (0.5 g) added to each container.	Sediment: 50 mL of acid washed sand

Table 3. Comparisons of the measurement endpoints of the SAM and MFC multispecies toxicity tests. Essentially the same levels of biological organization are included in both methods. In the calculation of clusters, derived variables are not particularly useful since they disproportionately weight certain measurements.

Measurement Endpoints

Standardized Aquatic Microcosm	Mixed Flask Culture
Primary Variables	**Primary Variables**
Population densities of inoculated organisms	Population densities of introduced organisms (often by classes such as diatoms, bluegreen bacteria, ostracods, protozoa etc.)
pH	pH
Photosynthesis/Respiration ratio	Photosynthesis/Respiration ratio
Optical Density	Optical Density
Analytical Chemistry of toxicant	Analytical Chemistry of toxicant
Nutrients	Nutrients
Bacterial counts	Bacterial counts
Derived variables	**Derived variables**
Algal Diversity	Algal Diversity
Total Algae	Total Algae
Available Algae	Available Algae
Total Daphnia	Total Daphnia
Total Invertebrates	Total Invertebrates

Table 4. Data analysis of the SAM and MFC multispecies toxicity tests. In our analyses, each system is analyzed using the same suite of statistical and artificial intelligence tools.

Data Analysis

Standardized Aquatic Microcosm	Mixed Flask Culture
Graphical Analysis (Plot the data) **Intervals of Non-significant Differences (IND)** **Metric Multivariate Statistics** **Non-metric multivariate Statistics-Riffle** **Projections-Space-time worms**	
Fewer species allow better identification and understanding of the potential role of each in the observed dynamics.	Thousands of species, and counting is often done at a variety of taxonomic levels. Not as much information on each of the organisms makes it difficult to assign roles and understand interactions. Many rare species

Comparison of Patterns in the SAM and MFC Test Results

The two methodologies have quite contrasting means of introducing organisms to the systems, and the operational volumes and surface to volume ratios are quite different. One manifestation of these differences is likely in the erratic nature of the clustering of the MFC compared to the SAM experiments conducted with the same toxicant. In Figure 2a, the occurrence of significant clustering in regards to treatment group follows a distinctive pattern for the SAM experiment, an initial significant clustering followed by a convergence of the treatments and then a re-emergence of the clustering. The MFC experiment reflects a much noisier pattern, one that calls into question whether or not the observed significant clustering is an artifact. Figure 2b compares the results of the SAM and MFC experiments. Note the noise inherent in the MFC as compared to the SAM system reflected in the NMCAA results.

In spite of the noise, and especially in the Jet-A experiments, an early and late period where the treatment groups are distinguishable seem to exist. In both sets of experimental protocols, Jet-A would have been seen to have generated more of an impact compared to JP-4, judging by the occurrence of significant clustering related to treatment effect.

As judged by the NMCAA results, none of the test systems demonstrated a recovery toward a stable system. This lack of recovery is reflected in both the significance of the clustering relative to treatment and the changing in the important variable rankings over sampling days. As an example, compare the last 3 sampling days for the JP-4 MFC experiment. The only variable deemed as important on all three days is "optical density". "Other Bluegreens", "Ostracod 2", and "P/R" are found on two of the sampling dates. The variables pH, *P. bursaria* and Nitzschia are found on only one sampling date each. The rapidly changing significance values found in both MFC tests also indicate a dynamic and rapidly evolving system.

Generic Multispecies Toxicity Tests

Microcosm testing strategies provide a greater dimensionality to toxicity testing, and resolve impacts that can not be extrapolated from single species toxicity tests. The MFC and the SAM do not try to simulate specific natural ecosystems, but they do utilize organisms having distinct interspecific and intraspecific interrelations and responses typical of natural environments. These methods also display many of the structural and functional properties of ecosystems, e.g., photosynthetic production/respiration dynamics, competition and succession, grazing effects, and nutrient cycling (Giddings, 1983; Suter, 1993; Taub, 1984). Microbial process are present and degradation of xenobiotics and the potential impacts of degradation products can be studied (Landis et al., 1993c).

The other main advantages of using these generic microcosms is that they are standardized in terms of species composition (Giddings, 1983; Suter, 1993). The importance of this simplicity and replicability in construction is that it allows closer examination of specific relationships and interactions in determining responses to direct and indirect effects, it reduces the dynamic heterogeneity that could potentially diffuse or hide effect responses, and it allows the comparison of results obtained in different laboratories (Suter, 1993).

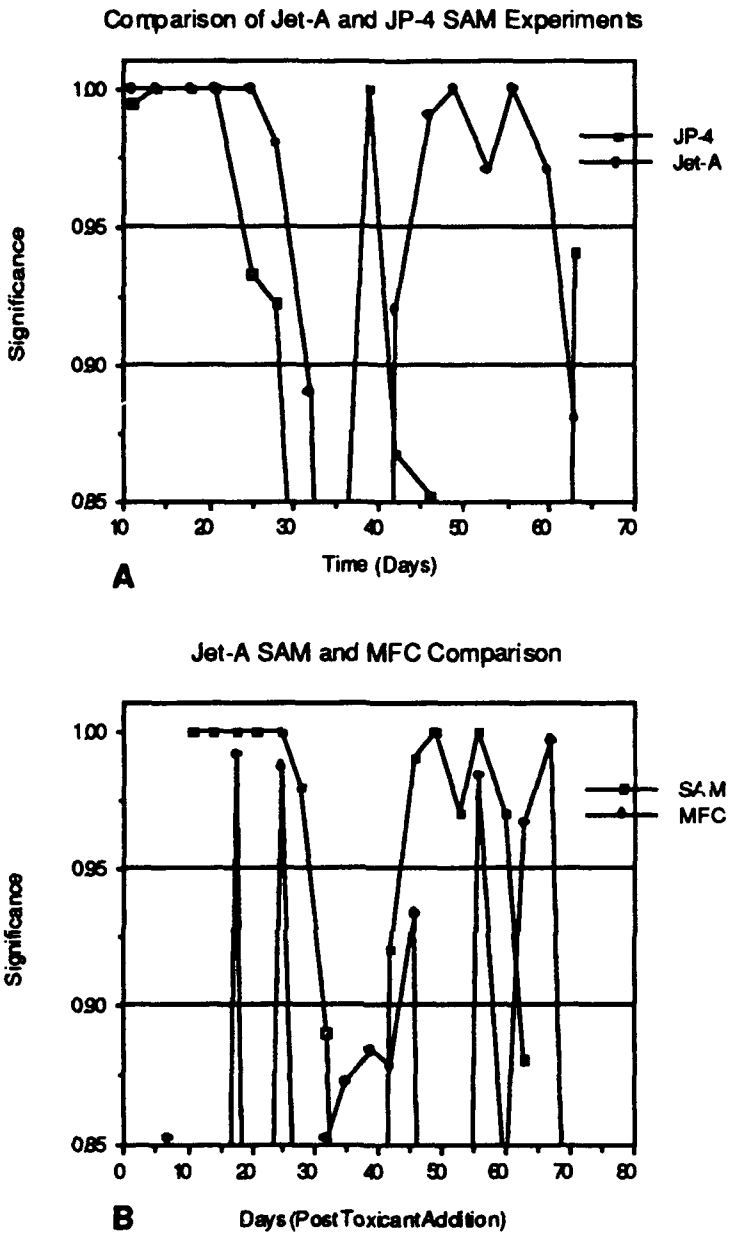


Figure 2. A comparison of the SAM and MFC NMCAA results for the Jet-A and JP-4.

The comparability and replicability of construction of a generic system is also a weakness. Since environmental heterogeneity, migration, colonization and other population, metapopulation and community level interactions are not modeled well in these systems, effects of toxicants upon these parameters will be difficult to ascertain. Numerous species representative of aquatic systems are not included, for example, fish and macrobenthos, and these organisms would be difficult to incorporate given the small size of the system.

Design Suggestions for Multispecies Toxicity Tests

The comparison of the two methods described here, along with our previous experience with microcosms and data analysis of these systems (Landis et al., 1993a; 1993b; 1993c; Matthews et al., in press; Haley et al., 1988; Matthews and Matthews, 1990; Sandberg 1994) leads us to suggest several improvements for the performance of multispecies toxicity tests. In several instances the suggestions are specific to the MFC and SAM systems, however many can be applied to systems regardless of size.

One of the most important aspects of any multispecies toxicity test is the realization by the investigators that these systems are models, inherently much more complex than computer simulations, of naturally occurring ecological systems. As has been demonstrated for the lakes studied by Katz et al. (1987), the best predictor of the future behavior of a system is itself. All model ecosystems will be limited in their predictive power, however, a primary advantage of model systems is that they are likely to also include interactions, parameters and relationships that are currently unknown and therefore impossible to simulate in an explanation based system. Because of this fact, multispecies toxicity tests are powerful tools in the investigation and eventual understanding of toxicant impacts in naturally occurring systems. Our suggestions are made in this light.

Parameter Selection, Measurement and Sampling Frequency

In both microcosm protocols, the parameters measured and the analyses conducted focus primarily on the biological structural components, including a few physical parameters, e.g., pH, dissolved oxygen, conductivity, and alkalinity. Species are identified and enumerated during the course of the experiment, to determine changes in diversity and abundance patterns. An important consideration is that these parameters are easily measured given the limited volumes and manpower requirements of performing the SAM or MFC tests. The premise of using this approach is that focusing on the functions, interactions, and responses of the individual parts will reveal ecosystem level dynamics (O'Neill and Waide, 1981). Each population variable can serve as an axis to track the movement of the system through ecosystem space. This approach is not without theoretical support. Ecosystems as perceived by the organisms are multidimensional. The Hutchinsonian idea of organisms and populations residing in a n-dimensional hypervolume is the basis of current niche theory (Hutchinson, 1959). The n-dimensional niche hypervolume is the ecosystem with all its components as perceived by the population. The variability of these parameters over time as well is used to account for the variety of species within the ecosystem (Hutchinson, 1961; Richerson et al., 1970; Tilman, 1982).

Other parameters should also be sampled, if possible, to increase the resolution of the toxicity tests. There are limitations to the using of components to assess effects to the whole ecosystem. Microbial processes often dominate the metabolism of aquatic systems, yet procaryotic populations are difficult to measure and their rapid turnover times makes frequent sampling necessary. Since a 24 hr period can be as many as 48 generations in procaryote populations, sampling on the scale of hours would be necessary. Although the population structure of filter feeding organisms can give an indication of the procaryotic assemblage, other parameters can give a more direct indication of the status of the procaryotic community.

Among these parameters are productivity/respiration ratios; total CO_2 efflux; biochemical rates; nutrient cycling; dissolved oxygen concentrations; pH; substrate decomposition rates; toxicant degradation rates; and accumulation rates of metabolic by-products (O'Neill and Waide, .981; Sugiura, 1992).

Cross Inoculation

The purpose of cross inoculation among replicate systems is generally seen as a means of ensuring the homogeneity of the test systems prior to treatment. However, this principally sets each replicate as an island with frequent migration that will maintain each system with a larger number of species than normal for that particular island size. Species that would normally become extinct are re-supplied in the inoculum. Upon the elimination of the cross inoculation followed by the toxicant addition, two factors are operating. First, a reduction in species as rare organisms become extinct. Second, the effects of the toxicant begin to operate. In effect, each of the 24 replicates starts from a different location in ecological space, no control can be exercised to force them into similarity, and finally a toxicant impacts the system. Cross inoculation seems to unduly complicate the methodology without an increase in sensitivity.

Impact of Multivariate Analysis and Community Conditioning on Risk Assessment and Environmental Restoration

Search for Relevant Assessment and Measurement Endpoints

Our current research indicates that identity of the variables that contribute the most to separating control treatment from dosed treatment groups change from sampling period to sampling period. The variables change in the SAM experiments, no doubt, in response to the successional trajectory of the system as nutrients become depleted. As nutrients become limiting and the ability of the system to exhibit large differences in community structure become less, the metric measures do not exhibit the same magnitudes of separation. Nonmetric clustering does not seem to be as sensitive to these changes.

However, the search for diagnostic measures to indicate the displacement of an ecosystem may not be fruitless. Although the relative importance of the variables in the SAM experiments may change, there are often variables that are more critical during the earlier stages of the development of the microcosm and those that are more crucial in the latter stages. The variable Ostracods is generally more important in the latter half of the experimental series than in the latter stages. The crucial aspect is that the clustering algorithm is able to select ecosystem attributes that are the best in differentiating stressed versus non-stressed systems. Although expert judgment may be able to predict in some cases variables that could be considered important to measure, the clustering approach is rapid, consistent, and not biased.

Instead of defining Assessment Endpoints, it may be more practical to define an Assessment Baseline or hypervolume using variables that have been demonstrated to be important in past descriptions of these types of ecosystems. Defining the 95 percent confidence region may be a more accurate way of characterizing the problem than by using artificial constructs or individual assessment measurement endpoint combinations. Assignment of these confidence regions may also improve the quality and accuracy of environmental risk assessment. Another logical outcome is that these regions must be defined by the measurement endpoints

(variables). Measurement endpoints are the means by which a system can be accurately placed and its trajectory defined in an n-dimensional coordinate system. Such a means of describing systems has already been proposed by Kersting (1988). The confidence region used to calculate NES is static, but an accounting of the passage of such a system through the coordinate system should provide a region from which deviation can be measured. Comparing dosed treatment groups to a control group is essentially the corresponding exercise but using a control series of replicates instead of an *a priori* prediction to measure deviation from the Assessment Baseline hypervolumes.

Measurement endpoints are therefore operationally defined as the variables that set the axes for the description of the system within the n-dimensional space. Data such as dose-response curves may play a part if they describe a relevant axes when used in a biomonitoring role. Dose response data, however, are not measurement endpoints by themselves, but are important in setting relevant system parameters. It is preferable to select measurement endpoints that are the lowest common denominator of the system that is capable of being measured. For example, pH is certainly the most direct measurement of hydrogen ion concentration available. Diversity and other indices of species number and community structure, however, are composites of species abundance data.

The Myth of Ecosystem Health and Measurement Indices

The use of indices such as diversity and the Index of Biological Integrity have the effect of collapsing the dimensions of the hypervolume in a relatively arbitrary fashion. Indices, since they are composited variables, are not true endpoints. The collapse of the dimensions that are composited to one tends to eliminate crucial information, such as the variability and distribution of the organisms within a particular system. The mere presence of absence and the frequency of these events can be analyzed using techniques such as nonmetric clustering and preserves the nature of the dataset. A useful function was certainly served by the application of these methods, but the new methods of data analysis and compilation should serve to replace these approaches and preserve the underlying structure and dynamic nature of ecological systems.

Part of the attraction of using indices may result in the pervasive nature of the metaphor, ecosystem health. In a recent critical evaluation, Suter (1993) dismissed ecosystem health as a misrepresentation of ecological science. Ecosystems are not organisms with the patterns of homeostasis determined by a central genetic core. Since ecosystems are not organismal in nature, health is a property that can not describe the state of such a system. The urge to represent such a state as health has lead to the compilation of variables with different metrics, characteristics and casual relationships. Suter suggests a better alternative would be to evaluate the array of ecosystem processes of interest, a process that is now possible given multivariate methods.

The Assumption of Non-equilibrium Dynamics in the Evaluation of Ecosystem Responses to Stressors

A common assumption in environmental toxicology is that after the initial stress, ecosystems recover to resemble the control state or reference site. These assumptions may be based more on outmoded theory than reality. Recent findings of complex dynamics in relatively simple microcosms, chaotic dynamics in

ecological field studies, and techniques of examining complex datasets demonstrate that non-equilibrium systems are the rule.

The use of nonmetric clustering in the analysis of ecological datasets has led us to formulate a non-equilibrium theory, the community conditioning hypothesis. The community conditioning hypothesis states that ecological communities preserve information about every event in their etiology. In our studies of standardized aquatic microcosms (SAMs), for example, we observed distinct community changes in response to stress that would appear and disappear over a two-month period (Landis et al., 1993a; Landis et al., 1993b). Even after the dosed systems had "recovered" to a state indistinguishable from the reference systems, a stress effect reappeared. A purely stochastic system could not exhibit this effect, since information is erased over time and two systems with identical distributions will remain identically distributed. A chaotic system could exhibit this effect, but we do not believe these microcosms are inherently chaotic, since similar systems tend to follow similar evolutions, without the divergences characteristic of nonlinear systems. Instead, we advanced the hypothesis that an unobserved feature of the community carried information about the stressor throughout the history of the system. In the case of the SAMs, we hypothesize detrital conditioning as the mechanism by which information is preserved. The preservation of the information can be contained in a variety of structural components of the ecological system, including genetics, competitive interactions, migration dynamics, community structure or age structure of a population. Examples of such conditioning can be found in the mitochondrial sequences of human populations and the affinity of 2,3,7,8 dioxin for the vertebrate Arh receptor.

An outgrowth of this research has been the development of a specific theory, that of Community Conditioning, that generates specific and testable hypotheses regarding ecosystems at the community level. As described above specific hypotheses generated from Community Conditioning include:

1. Biological communities may have long latency periods between observable cause and effects.
2. Communities may have patterns in common despite differences in compositions and physical scales. These patterns are more likely to be those at the system level rather than population level.
3. The history of an ecosystem is essential in determining the etiology of an effect.

Each of these factors effect the assessment of ecological risk.

Degradation of the WSF Components of Turbine Fuel in Microcosm Systems - A. Markiewicz.

Ms. Markiewicz has conducted an extensive investigation into the fate of the water soluble components of Jet-A and JP-8 in the MFC and SAM systems. The results are only summarized in this section, the thesis is included as part of Appendix D.

Degradation rates and biodegradation products of WSF from the fuels were monitored to evaluate whether the functional dynamics of the systems were similar regardless of the species structure and trophic complexity. The analysis was conducted using purge and trap gas chromatography with samples taken from the same microcosms being used for the community analyses. After the normal course of the microcosm experiment, microcosms from Treatment 1 (0 percent WSF) and Treatment 4 (15 percent WSF) were redosed

to 15 percent WSF to determine whether the degradation rates would be increased due to selective adaptation of microbial populations.

As can be seen in Figure 3, the WSF of both Jet-A and JP-4 are complex mixtures of materials with numerous peaks. However, within 48 hours post application, the concentrations are substantially reduced due to volatilization and physical and biological degradation (Figure 4). Further examination of the chromatograms also reveals that the WSF of the jet fuels are comprised of different concentrations and types of constituents, although they are similar cuts of the refining of petroleum.

In several cases, specific compounds were followed within each of the WSF fractions. Figure 5 depicts the concentration curves for benzene and toluene. After 48 hours both toluene and benzene are approximately one-half of their original concentration. At 192 hours (8 days) the concentrations are extremely low. Interestingly, at 192 hours for benzene and 240 hours for toluene, the materials reappear in the water column. This may be due to resuspension after release from dead and decaying organisms, or their appearance as by products of the degradation of higher molecular weight materials. These purges of materials into the water column occur at irregular intervals throughout the remainder of the experiment.

Further investigation demonstrated that the concentration of the class of hydrocarbons in the WSF determines the degradation rates rather than the concentration of a specific compound. This finding suggests that degradation pathways are generic regarding chemical class. Both microcosm types (MFC and SAM) displayed similar patterns of degradation and metabolite production dynamics. However, only the SAM displayed increased rates of hydrocarbon degradation in the retreated microcosms.

Multivariate Analysis of the Effects of a Pulsed Release of Jet-A Turbine Fuel from Sediments Using A Modified Mixed Flask Culture (MFC) Microcosm - R. Sandberg

The aquatic toxicity information used to satisfy regulatory requirements under FIFRA are generated under a tiered testing sequence with nearly all decisions regarding registration based on the results of single species tests. Over the last 15 years, a variety of multispecies aquatic toxicity tests have been developed with the hope that the increased complexity of the test system would result in a more realistic, community-level response to contamination. Sediments are often times a major repository for contaminants introduced into surface waters. The science of sediment toxicology itself, however, has been described as being in its infancy due to the failure to incorporate ecosystem disturbance into toxicity assessments.

This study investigates both the methods and the ecosystem level effects of producing a simulated release of a complex hydrocarbon mixture from sediments using a 60-day one liter modified Mixed Flask Culture (MFC) microcosm. A slow pulsed release of the test material from the spiked layer was obtained resulting in an initial period of perturbation caused by the transfer perturbation of the spiking procedure, as well as the effects of the hydrocarbon mixture. Monitored community structural parameters indicated that initial replicability was not obtained, the spatial scale of the MFC may be inadequate, and that treatment effects were generally detectable throughout the entire test with no apparent recovery or stability of the system. Multivariate techniques were able to distinguish statistically significant responses of the system holistically and reveal patterns not apparent with univariate results.

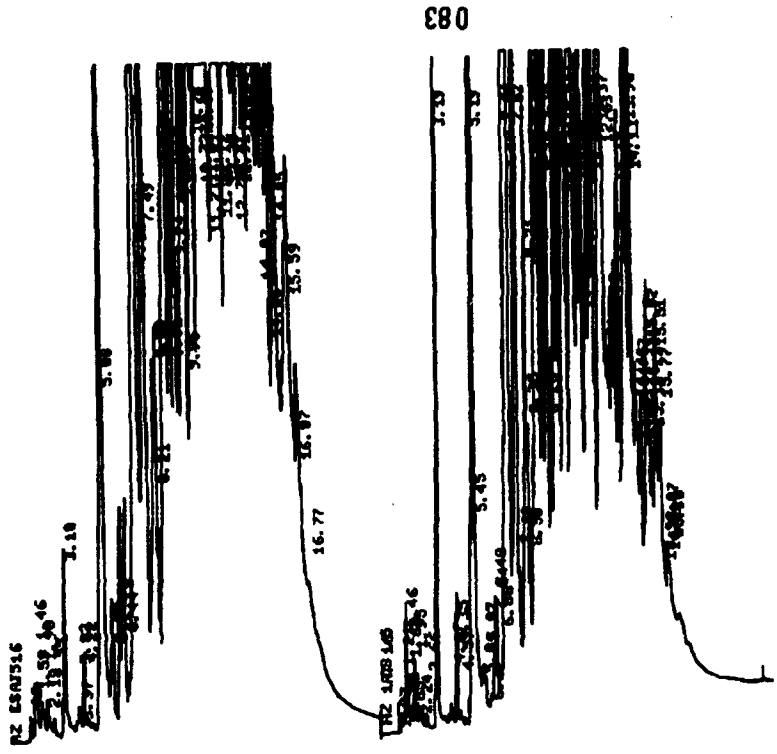


Figure 3. Gas Chromatographs of JP-8 15% WSF (left) and JET-A 15% WSF (right), immediately after treatment (0.0 hours).

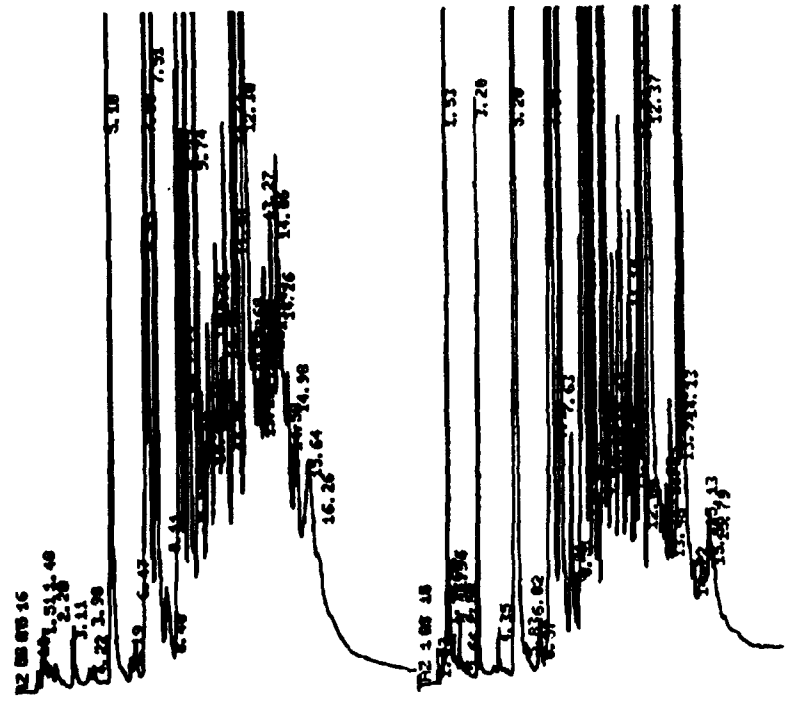


Figure 4. Gas chromatographs of JP-8 15% WSF (left) and JET-A 15% WSF (right), approximately 48 hours after treatment.

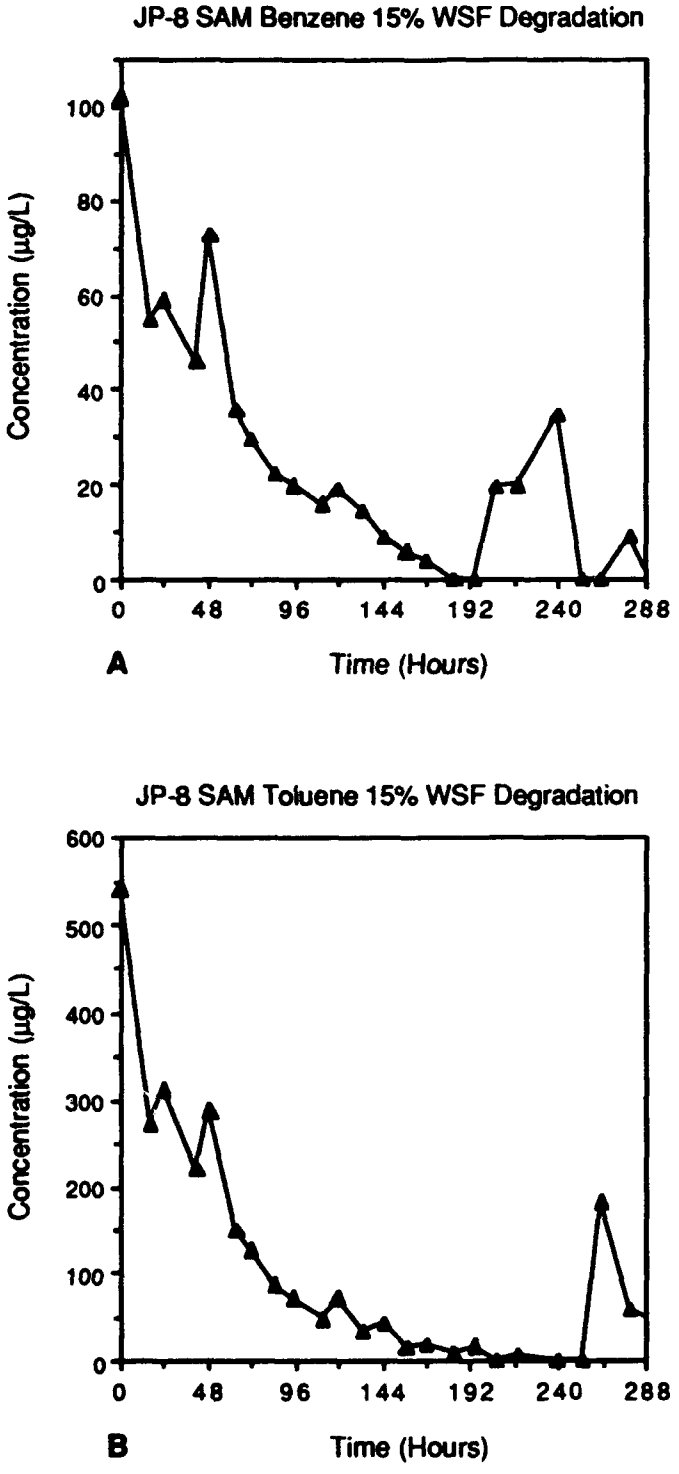


Figure 5. Fate curves of two of the constituents of the WSF of JP-8, Benzene and Toluene. Notice the rapid elimination of these materials from the water column, followed by a subsequent increase.

References

Bartell, S.M., R.H. Gardner and R.V. O'Neill. 1992. *Ecological Risk Estimation.* Lewis Publishers, Boca Raton, 1992, pp. 252.

Cheeseman, P., J. Kelly, M. Self, J. Stutz, W. Taylor and D. Freeman. 1988. Auto-class: A bayesian classification system. In Proceedings of the Fifth International Conference on Machine Learning, Los Altos, California. Morgan Kaufmann.

Domeshek, E.A., M.F. Herndon, A.W. Bennett and J.L. Kolodner. 1994. A case-based design aid for conceptual design of aircraft subsystems. In Proceedings of the Tenth Conference on Artificial Intelligence for Applications, San Antonio, TX. IEEE Computer Society Press.

Giddings, J.M. 1983. Microcosms for assessment of chemical effects on the properties of aquatic ecosystems. In J. Saxena, ed, Hazard Assessment of Chemicals. Current Developments, Vol. 2., Academic Press, Inc., New York, pp. 45-94.

Haley, M.V., D.W. Johnson and W.G. Landis. 1988. The aquatic toxicity of brass dust. In *Aquatic Toxicology and Environmental Fate: Tenth Volume ASTM STP -971.* W. Adams, G. Chapman and W.G. Landis, eds., American Society for Testing and Materials, Philadelphia, pp. 468-479.

Hutchinson, G.E. 1959. Concluding remarks. *Cold Spring Harbor Symposium on Quanitative Biology.* **22**:415-427.

Hutchinson, G.E. 1961. The paradox of the plankton. *Amer. Nat.* **95**:137-143.

Katz, T.K., T.M. Frost and J.J. Magnuson. 1987. Inferences from spatial and temporal variability in ecosystems: Long-term zooplankton data from lakes. *Amer. Nat.* **129**:830-846.

Kauffman, S.A. and S.D. Johnsen. 1991. Coevolution to the edge of chaos; coupled fitness landscapes, poised states and coevolutionary avalanches. *J. theor. Biol.* **149**:467-505.

Kauffman, S.A. 1993. *The Origins of Order, Self-Organization and Selection in Evolution.* Oxford University Press Inc., New York, New York.

Kersting, K. 1988. Normalized ecosystem strain in micro-ecosystems using different sets of state variables. *Verh. Internat. Verein. Limnol.* **23**:1641-1646.

Landis, W.G., R.A. Matthews, A.J. Markiewicz, N.A. Shough and G.B. Matthews. 1993a. Multivariate Analyses of the Impacts of the Turbine Fuel Jet-A Using a Microcosm Toxicity Test. *J. Environ. Sci.* Vol 2:113-130.

Landis, W.G., R.A. Matthews, A.J. Markiewicz and G.B. Matthews. 1993b. Multivariate Analysis of the Impacts of the Turbine Fuel JP-4 in a Microcosm Toxicity Test with Implications for the Evaluation of Ecosystem Dynamics and Risk Assessment. *Ecotoxicology* 2:271-300.

Landis, W.G., M.V. Haley and N.A. Chester. 1993c. The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In W. G. Landis, J. Hughes and M. Lewis, *Environmental Toxicology and Risk Assessment ASTM STP -1167.* American Society for Testing and Materials, Philadelphia. pp. 159-177

Landis, W.G., G.B. Matthews, R.A. Matthews and A. Sergeant. *In press, 1994.* Application of multivariate techniques to endpoint determination, selection and evaluation in ecological risk assessment. *Environ. Toxicol. Chem.*

Lebowitz, M. 1990. The utility of similarity-based learning in a world needing explanation. In Kodratoff, Y. and Michalski, R. S., editors, Machine Learning, An Artificial Intelligence Approach, Volume III, pp. 399--422. Morgan Kaufmann, Los Altos, California.

Markiewicz, A.J. 1994. Comparison of the Degradation of Water Soluble Components in Jet Fuel Using the Standard Aquatic Microcosm (SAM) and Mixed Flask Culture (MFC). M.S. Thesis, Western Washington University.

Matthews, G. and J. Hearne. 1991. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175-184.

Matthews, G. and R. Matthews. 1990. A model for describing community change. In Pesticides in Natural Systems: How Can Their Effects Be Monitored? Proceedings of the Conference, Environmental Research Laboratory/ORD, Corvallis Oregon. EPA 9109/9-91-011.

Matthews, G.B., R.A. Matthews and W.G. Landis. *In press.* Nonmetric clustering and association analysis: Implications for the evaluation of multispecies toxicity tests and field monitoring. *Environmental Toxicology and Risk Assessment-Third Volume, ASTM 1218,* J.S. Hughes, G.R. Biddinger and E. Mones, eds., American Society for Testing and Materials, Philadelphia, 1994.

Michalski, R.S. and R.L. Chilausky. 1980. Learning by being told and learning from examples: an experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean diseases. Policy Analysis and Information Systems, 4.

Nicolis, G. and I. Prigogine. 1989. *Exploring Complexity.* W.H. Freeman and Co. New York.

O'Neill, R.V. and J.B. Waide. 1981. Ecosystem theory and the unexpected: Implications for environmental toxicology. In B. W. Cornby, ed., Management of Toxic Substances in Our Ecosystems. Taming the Medusa. Ann Arbor Science Publishers, Inc., pp. 43-73.

Richerson, P., R. Armstrong and C.R. Goldman. 1970. Contemporaneous disequilibrium, a new hypothesis to explain the "paradox of the plankton". *Proc. Natl. Acad. Sci. U.S.A.* **67**:1710-1714.

Sandberg, R. 1994. Investigation of the Effects of a Pulsed Release of Jet-A Turbine Fuel From Sediments Using a Modified Mixed Flask Culture (MFC) Microcosm. M.S. Thesis, Western Washington University.

Sugiura, K. 1992. A multispecies laboratory microcosm for screening ecotoxicological impacts of chemicals. *Env. Tox. Chem.* **11**:1217-1226.

Suter, G.W. II. 1993. A critique of ecosystem health concepts and indexes. *Environ. Toxicol. Chem.* **12**:1533-1539.

Taub, F.B. 1984. Synthetic microcosms as biological models of algal communities. In L. E. Shubert, ed., Algae as Ecological Indicators. Academic Press, Inc., London, pp. 363-394.

Tilman, D. 1982. *Resource Competition and Community Structure.* Princeton University Press, Princeton, NJ, pp. 11-138.

Publications

Journals:

Landis, W.G., M.V. Haley and N.A. Chester. 1993 The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In W. G. Landis, J. Hughes and M. Lewis, *Environmental Toxicology and Risk Assessment ASTM STP -1167.* American Society for Testing and Materials, Philadelphia, pp. 159-177.

Landis, W. G. and J. Hughes and M. Lewis. 1993. *Environmental Toxicology and Risk Assessment: ASTM STP -1179.* American Society for Testing and Materials, Philadelphia.

Landis, W. G., R. A. Matthews, A.J. Markiewicz, N. A. Shough and G. B. Matthews. 1993. Multivariate Analyses of the Impacts of the Turbine Fuel Jet-A Using a Microcosm Toxicity Test. *J. Environ. Sci.* Vol 2:113-130.

Landis, W. G., R. A. Matthews, A. J. Markiewicz and G. B. Matthews. 1993. Multivariate Analysis of the Impacts of the Turbine Fuel JP-4 in a Microcosm Toxicity Test with Implications for the Evaluation of Ecosystem Dynamics and Risk Assessment. *Ecotoxicology* 2:271-300.

Landis, W.G., R.A. Matthews and G.B. Matthews. 1993. Complexity, irreversibility and multispecies systems. *SETAC News* 13:5 pp 8-9.

In Press 1994:

Matthews, G. B., R. A. Matthews and W. G. Landis. *In press.* Nonmetric Conceptual Clustering in Ecology and Ecotoxicology. *AI Applications.*

Landis, W. G., G. B. Matthews, R. A. Matthews and A. Sergeant. *In press.* Application of multivariate techniques to endpoint determination, selection and evaluation in ecological risk assessment. *Environ. Toxicol. Chem.*

Landis, W. G., Matthews, R. A., Markiewicz, A. J. and Matthews, G. B. *In press.* Non-linear oscillations detected by multivariate analysis in microcosm toxicity tests with complex toxicants: Implications for biomonitoring and risk assessment. *Environmental Toxicology and Risk Assessment-Third Volume, ASTM 1218,* Jane S. Hughes, Gregory R. Biddinger, and Eugene Mones, Eds., American Society for Testing and Materials, Philadelphia, 1994.

Matthews, G. B., R. A. Matthews and W. G. Landis. *In press.*. Nonmetric clustering and association analysis: Implications for the evaluation of multispecies toxicity tests and field monitoring. *Environmental Toxicology and Risk Assessment-Third Volume, ASTM 1218,* Jane S. Hughes, Gregory R. Biddinger, and Eugene Mones, Eds., American Society for Testing and Materials, Philadelphia, 1994.

Landis, W. G. Book Review. *In press.* Ecological Risk Analysis by S. M. Bartell, R. H. Gardner and R. V. O'Neill. *Aquatic Toxicology.*

Submitted:

Landis, W. G., A. J. Markiewicz, R. A. Matthews, M. J. Roze and G. B. Matthews. Comparison of the community dynamics of the standardized aquatic microcosm and mixed flask culture microcosm toxicity tests with suggestions for design criteria in multispecies toxicity testing. in *Ecological Risk Assessment of Pesticides: Enhancing the Process. SETAC Special Publication.*

Masters Thesis Completed

Sandberg, R.S. 1994. Investigation of the Effects of a Pulsed Release of Jet-A Turbine Fuel from Sediments using a Modified Mixed Flask Culture (MFC) Microcosm. M. S. Thesis in Environmental Science, Western Washington University, Bellingham.

Keel, L.W. 1994. Sublethal Toxicity of Copper Sulfate to the Intertidal Sea Anemone, *Anthopleura elegantissima.* M. S. Thesis in Environmental Science, Western Washington University, Bellingham.

Markiewicz, A.J. 1994. Comparison of the Biodegradation of Water Soluble Components in Jet Fuel Using the Standardized Aquatic Microcosm (SAM) and the Mixed Flask Culture Microcosm (MFC). M. S. Thesis in Environmental Science, Western Washington University, Bellingham.

Presentations, 1991-1994

Landis, W.G. Environmental risk assessment framework incorporating ecosystem level factors. Annual Meeting of the Society of Environmental Toxicology and Chemistry, Seattle, WA, November 5, 1991.

Landis, W.G., G.B. Matthews, R.A. Matthews, and N.J. Shough. Evaluation of the Aquatic Toxicity of the Turbine Fuel Jet-A Using Single Species and Microcosm Toxicity Tests. ASTM Environmental Toxicology and Risk Assessment Symposium, Pittsburgh, PA, April 27, 1992.

G.B. Matthews, R.A. Matthews, and W.G. Landis. Multivariate Analyses of Data from Aquatic Toxicity Microcosm Studies: a Comparison of Three Statistical Tests. ASTM Environmental Toxicology and Risk Assessment Symposium, Pittsburgh, PA, April 27, 1992.

Landis, W.G., G.B. Matthews and R.A. Matthews. The use of multispecies toxicity tests and their role in environmental risk assessment. School of Fisheries seminar, University of Washington, Seattle, WA, May 21, 1992.

Matthews, G.B., R.A. Matthews and W.G. Landis. Multivariate Analyses of Data from Aquatic Toxicity Microcosm Studies: A Comparison of Three Statistical Tests Pacific Northwest SETAC Annual Meeting, Bellingham, WA, June 26-27, 1992.

Landis, W.G., G.B. Matthews, R.A. Matthews and N.J. Shough. Evaluation of the Aquatic Toxicity of the Turbine Fuel Jet-A using Single Species and Microcosm Toxicity Tests. Pacific Northwest SETAC Annual Meeting, Bellingham, WA, June 26-27, 1992.

Keel, L.W. and W.G. Landis. Sublethal Toxicity of Copper to the Sea Anemone *Anthropleura elegantissima*. Pacific Northwest SETAC Annual Meeting, Bellingham, WA, June 26-27, 1992.

Landis, W.G., R. Sahakian, F.B. Taub and J.H. Taub. Population effects and community dynamics in the standardized aquatic microcosm-modeling compared to historical data sets. Annual Meeting Society of Environmental Toxicology and Chemistry, Cincinnati, OH, November 1992.

Landis, W.G. , R.A. Matthews, A.J. Markiewicz, N.J. Shough and G.B. Matthews. Evaluation of the aquatic toxicity of two turbine fuels using microcosm toxicity tests. Annual Meeting Society of Environmental Toxicology and Chemistry, Cincinnati, OH, November 1992.

Matthews, G.B., W.G. Landis, and R.A. Matthews. Replicability of the control group response in the SAM. Annual Meeting Society of Environmental Toxicology and Chemistry, Cincinnati, OH, November 1992.

Landis, W.G. and A. Sergeant. Incorporation of multivariate techniques to the performance of ecological risk assessments. Annual Meeting Society of Environmental Toxicology and Chemistry, Cincinnati, OH, November 1992.

Landis, W.G., R.A. Matthews and G.B. Matthews. Endpoints: art or artifact? Annual Meeting Society of Environmental Toxicology and Chemistry, Cincinnati, OH, November 1992.

Matthews, G.B., W.G. Landis and R.A. Matthews. Nonmetric clustering and association analysis: an artificial intelligence approach to multispecies data. Annual Meeting Society of Environmental Toxicology and Chemistry, Cincinnati, OH, November 1992.

Landis, W.G., R.A. Matthews and G.B. Matthews. Community level responses to toxicant stress as evaluated by nonmetric clustering and multivariate projections, implications for resource damage and risk assessment. U.S. EPA-ERL, Corvallis, OR, January 1993.

Landis, W.G., R.A. Matthews and G.B. Matthews. Ecosystem Dynamics: Wormspace, Chaos and the Implications for Ecological Risk Assessment. Keynote Address, Ecological Risk Assessment. USEPA Regional Risk Assessment Annual Meeting, Atlanta, GA, March 4, 1993.

Landis, W.G., R.A. Matthews and G.B. Matthews. Ecosystem Dynamics: Wormspace, Chaos and the Implications for Ecological Risk Assessment. Lecture, Classic Papers in Aquatic Science Course, University of Washington, March 6, 1993 Seattle, WA.

Landis, W.G., R.A. Matthews and G.B. Matthews. Ecosystem Dynamics: Wormspace, Chaos and the Implications for Ecological Risk Assessment. USEPA Region 10 Risk Assessment Seminar Series, Seattle, WA, March 24, 1993.

Landis, W.G., R.A. Matthews and G.B. Matthews. Oscillations detected by multivariate analysis in microcosm toxicity tests with complex toxicants: implications for risk assessment. Society of Environmental Toxicology and Chemistry World Congress, Lisbon, Portugal, March 30, 1993.

Matthews, G.B., W.G. Landis and R.A. Matthews. Nonmetric clustering and association analysis: risk assessment implications for the interpretation of multispecies toxicity tests and field monitoring. Society of Environmental Toxicology and Chemistry World Congress, Lisbon, Portugal, March 30, 1993.

Grue, C. and W.G. Landis. Indirect effects of contaminants on aquatic wildlife. Society of Environmental Toxicology and Chemistry World Congress, Lisbon, Portugal, March 30, 1993.

G. B. Matthews, W.G. Landis, and R. A. Matthews. Nonmetric Clustering and Association Analysis: Implications for the Evaluation of Multispecies Toxicity Tests and Field Monitoring, April 23, 1993.

Landis, W.G., R.A. Matthews and G.B. Matthews. Oscillations detected by multivariate analysis in microcosm toxicity tests with complex toxicants: implications for biomonitoring and risk assessment. ASTM Symposium Environmental Toxicology and Risk Assessment, Atlanta, GA, April 27, 1993.

Matthews, G.B., W.G. Landis and R.A. Matthews. Nonmetric clustering and association analysis: implications for the evaluation of multispecies toxicity tests and field monitoring. ASTM Symposium Environmental Toxicology and Risk Assessment, Atlanta, GA, April 27,1993.

Sandberg, R.S. and W.G. Landis. Use of the Mixed Flask Culture (MFC) Microcosm Protocol to Investigate the Effects of a Pulsed Release of Jet-A Turbine Fuel from Sediments. 1993 Annual Meeting Pacific Northwest Chapter of the Society of Environmental Toxicology and Chemistry, Newport OR, May 20-22, 1993.

Rodgers, S.C. and W.G. Landis. Evaluation of Community Structures Vs Community Function after Exposure to the Turbine Fuel Jet-A. 1993 Annual Meeting Pacific Northwest Chapter of the Society of Environmental Toxicology and Chemistry, Newport OR, May 20-22, 1993.

Markiewicz, A.J., R.A. Matthews and W.G. Landis. Comparison of the Degradation of Water Soluble Components in Jet Fuel using the Standardized Aquatic Microcosm (SAM) and the Mixed Flask Microcosm (MFC). 1993 Annual Meeting Pacific Northwest Chapter of the Society of Environmental Toxicology and Chemistry, Newport, OR, May 20-22, 1993.

Landis, W.G., A.J. Markiewicz, R.A. Matthews and G.B. Matthews. Non-linear Dynamics of Microcosm Experiments after Toxicant Stress Evaluated by Response Volume Projections (Space-Time Worms). 1993 Annual Meeting Pacific Northwest Chapter of the Society of Environmental Toxicology and Chemistry, Newport, OR, May 22, 1993.

Landis, W.G., R.A. Matthews and G.B. Matthews. Use of Novel Methods and the Application of Non-linear Dynamics to the Evaluation of Ecosystem Impacts. Seminar National Oceanic and Atmospheric Administration Hazardous Materials Team, Seattle, WA, June 21, 1993.

Matthews, G.B., R.A. Matthews and W.G. Landis. Nonmetric Clustering and Association Analysis and its Use in Hazard Assessment. SETAC Short Course, Houston, TX, November 1993.

Landis, W.G., R.A. Matthews, A.J. Markiewicz and G.B. Matthews. Comparison of Test Results in the Evaluation of the WSF of Several Jet Fuels Using the Standardized Aquatic Microcosm and the Mixed Flask Culture Protocols. Annual Meeting Society of Environmental Toxicology and Chemistry, Houston, TX, November 1993.

Matthews, G.B., W.G. Landis and R.A. Matthews. Response Volumes (Space-time Worms) as a Method for the Visualization of Ecosystem Dynamics and Indirect Effects. Annual Meeting Society of Environmental Toxicology and Chemistry, Houston, TX, November 1993.

Landis, W.G., R.A. Matthews and G.B. Matthews. Non-linear Dynamics of Microcosm Ecosystems and the Inherent Limitations of Risk Assessment. Annual Meeting Society of Environmental Toxicology and Chemistry, Houston, TX, November 1993.

Landis, W.G., R.A. Matthews and G.B. Matthews. Characterization and Classification of Direct and Indirect Effects at the Community and Ecosystem Levels. Annual Meeting Society of Environmental Toxicology and Chemistry, Houston, TX, November 1993.

Markiewicz, A.J., R.A. Matthews and W.G. Landis. Comparison of the Degradation of Water Soluble Components in Jet Fuel Using the Standard Aquatic Microcosm (SAM) and the Mixed Flask Microcosm (MFC). Annual Meeting Society of Environmental Toxicology and Chemistry, Houston, TX, November 1993.

Rodgers, S.C. and W.G. Landis. Evaluation of Community Structure and Community Function After Exposure to the Turbine Fuel Jet-A. Annual Meeting Society of Environmental Toxicology and Chemistry, Houston, TX, November 1993.

Sandberg, R.S., M.J. Roze and W.G. Landis. Use of the Mixed Flask Culture (MFC) Microcosm Protocol to Investigate the Effects of a Pulsed Release of Jet-A. Annual Meeting Society of Environmental Toxicology and Chemistry, Houston, TX, November 1993.

Landis, W.G. , R.A. Matthews and G.B. Matthews. Community Conditioning as an Alternative to the Stability and Recovery of Ecosystem Hypothesis in Ecological Risk Assessment. ASTM Symposium on Environmental Toxicology and Risk Assessment, Montreal, April 12, 1994.

Matthews, G.B., R.A. Matthews and W.G. Landis. Artificial Intelligence Based Data Analysis and Visualization Tools for Ecological Risk Assessment . ASTM Symposium on Environmental Toxicology and Risk Assessment, Montreal, April 12, 1994.

Landis, W.G., R.A. Matthews, A.J. Markiewicz and G.B. Matthews. Structural and Community Level Comparison of Turbine Fuel Test Results Using the Standardized Aquatic Microcosm (SAM) and the Mixed Flask Culture (MFC) Protocols. ASTM Symposium on Environmental Toxicology and Risk Assessment, Montreal, April 12, 1994.

Landis, W.G. R.A. Matthews, G.B. Matthews. The Recovery Myth and an Alternative-Community Conditioning. Annual Meeting of the Pacific Northwest Chapter of the Society for Environmental Toxicology and Chemistry, University of Victoria, Victoria, B.C., May 21, 1994.

Scheduled for summer 1994

Landis, W.G., R.A. Matthews, G.B. Matthews and A.M. Markiewicz. The Community Conditioning Hypothesis and the Dynamics of Stressed Ecological Systems. Ecological Society of America Workshop Relevancy of Ecological Data to Pesticide Registration, Knoxville, TN, August 7, 1994.

Submitted for SETAC 94

Landis, W.G., and R.A. Matthews, G.B. Matthews, The Inherent Limitations of Population Modeling in Environmental Risk Assessment and an Alternative: Community Conditioning.

Matthews, R.A. W.G. Landis and G.B. Matthews. Application of the Community Conditioning Hypothesis to the Design of Multispecies Toxicity Tests.

Landis, W.G., R.A. Matthews M.A. Roze and G.B. Matthews. The Stability Myth and the Dynamics and Patterns of Xenobiotic Impacts to Ecological Systems.

Landis, W.G. M.A. Roze, G.B. Matthews, S. Dominguez and A. Fairbrother. A Multivariate Artificial Intelligence Approach to the Evaluation of Biomarkers Under Field Conditions II.

Graduate Students Supported by the Grant and Student Research Projects

Keel, Lester - Anthopluera as a Monitor for the Environmental Impacts of Toxicants (Dr. Landis-Huxley College).

Markiewicz, April - Fate of Jet Fuel Water Soluble Fraction in the Standardized Aquatic Microcosm (Dr. R. Matthews-Huxley College)

Rodgers, Sara - Comparison of MFC toxicity tests with and without adapted communities (Dr. Landis-Huxley College).

Sahakian, Robert - Population Dynamics and the Effects of Toxicants on Community Structure (Dr. Landis-Huxley College).

Sandberg, Randy - Modification of the MFC for use in sediment testing (Dr. Landis-Huxley College).

Roze, Michael - Application of RIFFLE program for data evaluation (Dr. G. Matthews-Computer Science).

Professional Collaborators in the Research Program

James W. Hearne, Computer Science Department, Western Washington University

Anne Fairbrother, USEPA Environmental Research Laboratory, Corvallis, OR and ep&t, Corvallis, OR

Richard Bennett, USEPA Environmental Research Laboratory, Corvallis, OR

Anne Sergeant, Office of Research and Development, U.S. EPA, Washington, DC

Frieda B. Taub, School of Fisheries, University of Washington, Seattle, WA

John H. Taub, Seattle, WA.

Larry Kaputska, ep&t, Corvallis, OR

Marsha Black, Oklahoma State University, Stillwater, OK

Steve Domingus, USEPA Environmental Research Laboratory, Corvallis, OR

Interactions and Consultations

Over the last year this research has been translated into technology transfers to DOD and EPA laboratories and the private sector. Apart from presenting the research at national and international meetings, we have been successful in transferring this data and technology during informal meetings or presentations on-site. Below is a list of several of the groups with which we met and transferred information over the last 36 months.

Joseph Dulka, Agricultural Product Department, DuPont Experimental Station, Wilmington, DE. Microcosm use and data analysis.

Lidia Watrud, Team Leader, and Ray Siedler Biotechnology Team, U.S. EPA-Corvallis, OR. Data analysis from terrestrial microcosms.

Nigel Blakley, Department of Ecology, Olympia, WA. Toxicity evaluation of petroleum mixtures.

SETAC Microcosm Workshop. Design and data analysis of microcosms for pesticide evaluations.

ICI Americas. Data analysis of aquatic microcosm studies.

Anne Sergeant, ORD, U.S. EPA., Washington, D.C. Application of multivariate methods to ecological risk assessments.

Heather Gordon, National Research Council, Canada. Riffle program.

Joni A. Torsella, U.S. EPA, Cincinnati, OH. Riffle program.

Patrick A. Thorpe, Grand Valley State University, Allendale, MI. Permtest program.

Charles Hadden, Science Applications International Corp., Oak Ridge, TN. Clustering analysis.

Byron Bodo, Byron A. Bodo & Associates, Canada. Nonmetric clustering techniques and Riffle program.

Prof. Hein H. Du Preez, Rand Afrikaans University, South Africa. AI techniques for multispecies toxicity tests.

Scott Ferson, Applied Biomathematics, Setauket, NY. Nonmetric clustering techniques.

Technology Transfers

SETAC Short Course: **"Nonmetric Clustering and Association Analysis in Ecotoxicology"**, November 14, 1994, Geoffrey Matthews, Mike Roze, Robin Matthews, Wayne Landis.

NOAA Course on Multivariate Statistics: **"Statistical Ecology Minicourse"**, March 17-18, 1993, Taught by Drs. Robin and Geoffrey Matthews.

Evaluation of Molecular Marker Datasets: **"The Evaluation of Biomarkers Under Field Conditions"**, S. Dominguez[1], A. Fairbrother[2], T. Shiroyama[1], and P. Bucholz[3]. [1]USEPA, Corvallis, OR; [2]ecological planning and toxicology, inc., Corvallis, OR; [3]Computer Sciences Corporation, Corvallis, OR. The application of biomarkers to ecotoxicology has been limited by large interindividual variability caused by genetic differences and simultaneous exposure to multiple stressors. This study describes the use of a multivariate approach to analyzing biomarker data from pesticide field studies. Gray-tailed voles (*Microtus canicaudus*) were placed in each of 24 0.2-ha enclosures planted with alfalfa. Three months later, populations reached densities of approximately 60 voles per enclosure. Azinphos methyl (Guthion 2S) was applied using a boom sprayer at 0, 0.77, 1.55, 3.11 and 4.67 kg active ingredient/ha. Ten adult voles were live-trapped from each enclosure on days 2,3,4,14,15,16 post-spray, bled, and released at the trap station where they were captured. On days 6,7,8,10,11,12 ten voles were trapped in each of four enclosures, bled, and killed to remove brains for analysis of cholinesterase activity. The following were measured in each blood sample: hematocrit; total and differential leukocyte count, hemoglobin, blood urea nitrogen, creatinine, creatine phosphokinase, isocitrate dehydrogenase, and lactate dehydrogenase. Summary statistics will be presented demonstrating the large variation within groups and among days. Analysis of variance techniques showed no differences among mean values for each of the treatment groups. Brain cholinesterase activity was significantly depressed in voles from azinphos methyl enclosures.

The biomarker data were derived from field experiments using gray-tailed voles placed in 0.2-ha field enclosures and dosed with azinphos methyl. Molecular markers included brain cholinesterase activity, blood chemistry, enzymatic and cell type markers. Data were analyzed using nonmetric clustering and association analysis (NMCAA), an artificial intelligence technique. NMCAA confirmed the ANOVA results in that brain cholinesterase activity was an important variable in clustering on treatment group. However, NMCAA found

that neutrophils and basophils were also important variables. The alteration of the ratio of leukocyte types has been previously reported in laboratory tests with azinphos methyl. The quality measurements of the clustering suggested that additional patterns are present. Using quadrants within the field experiment as a treatment, a statistically significant relationship was again found. The variables determined to be important were brain cholinesterase inhibition, pregnancy and basophils. Controlling for quadrant effects, again a significant association between dose and clusters was found within the quadrants even though the total sample size within a quadrant was significantly reduced. The biomarkers and NMCAA detected at least two important patterns in the field experiment, dose and location.

Abstracts of Papers Presented June 1 ,1993-May 31, 1994

Abstracts of the 1993 Society of Environmental Toxicology and Chemistry Meeting, Houston, Texas
Comparison of Test Results in the Evaluation of the WSF of Several Jet Fuels Using the Standardized Aquatic Microcosm and the Mixed Flask Culture Protocols. W.G. Landis, Matthews, R.A., and Markiewicz, A.J., Institute of Environmental Toxicology and Chemistry, Huxley College; Matthews, G.B., Computer Science Department, Western Washington University, Bellingham, WA. The water soluble fraction of the turbine fuels Jet-A, JP-4 and JP-8 have been examined as stressors for two microcosm protocols, the standardized aquatic microcosm (SAM) and the mixed flask culture (MFC). The SAM is a 3 L system inoculated with standard cultures of algae, zooplankton, bacteria, and protozoa. In contrast, the MFC is 1 L and is inoculated with a complex mixture of organisms derived from a natural source. Analysis of the organism counts and physical data were conducted using conventional and newly derived multivariate methods. Physical parameters, such as pH and oxygen metabolism, were often not as sensitive as species and bacterial counts. Like the SAM system, species numbers and other variables that determined clusters varied among sampling dates. Compared to the larger yet simpler system, the MFC exhibits more violent dynamics and is more likely to become catastrophically fixated, as in systems dominated by cyanobacteria. The combination of greater diversity and smaller volume may contribute to the volatile or chaotic dynamics of the MFC system.
This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.
Key Words: microcosms, standard aquatic microcosm, mixed flask culture, chaos, complexity, ecosystems

Characterization and Classification of Direct and Indirect Effects at the Community and Ecosystem Levels. W.G. Landis and Matthews, R.A., Institute of Environmental Toxicology and Chemistry, Huxley College; Matthews, G.B., Computer Science Department, Western Washington University, Bellingham, WA. The dynamics of the response of an ecosystem to a stressor have classically been separated into direct and indirect effects. The initial direct effects of a toxicant alter the community in two ways. First, the system can be displaced from its initial state. The magnitude of the displacement may be estimated using current laboratory toxicity tests, however, given the complexity or even chaotic nature of ecosystems, the directional vector of this displacement may be impossible to predict. Second, the dispersion or variability of the system can also be altered. In some instances the variability of the system can be radically decreased or increased depending upon the type of toxicant. Indirect effects, however, may be so persistent as to take another stressor event to remove the impacts of this history from the system. In our studies, recovery in the classical sense of returning to the original or reference state is unlikely to occur. Even in unstressed systems small initial differences give rise to dramatic changes. The accurate prediction of direction and magnitude of the indirect effect may prove impossible if ecosystems exhibit sufficiently complex or chaotic dynamics.
This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.
Key Words: direct and indirect effects, chaos, complexity, ecosystems

Non-linear Dynamics of Microcosm Ecosystems and the Inherent Limitations of Risk Assessment. W.G. Landis and Matthews, R.A., Institute of Environmental Toxicology and Chemistry, Huxley College; Matthews, G.B., Computer Science Department, Western Washington University, Bellingham, WA. Projections into two dimensional space with time are used to visualize ecosystem dynamics. The space-time worm projections have demonstrated that the systems are moving in a complex dynamic that does not repeat or recover as defined as the return of the dosed system to the space and dynamics of the non-dosed case. In cases where the dosed and non-dosed treatments overlap, the subsequent dynamics demonstrated that it is a case of passing through and not recovery. The patterns appear to be chaotic, such as turbulence and weather. Ecological important properties of these systems are: they do not return to an

original condition upon perturbation; the history of the perturbation resets the initial conditions making a return to the initial state virtually impossible; history of the system is important in setting the potential dynamics; and that predictions are limited not by knowledge but by the inherent dynamics of the system. Risk assessments and projections of impacts upon populations and communities have inherent limits on their power of prediction. These limits are inherent to the underlying dynamics of the system and not based on the uncertainty of the available knowledge.
This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.
Key Words: chaos, risk assessment, non-linear dynamics, ecosystems

Response Volumes (Space-time Worms) as a Method for the Visualization of Ecosystem Dynamics and Indirect Effects. G.B. Matthews, Computer Science Department; Landis, W.G., and Matthews, R.A., Institute of Environmental Toxicology and Chemistry, Huxley College, Western Washington University, Bellingham, WA. A variety of indexes and other composite measures of ecosystems, such as measures of integrity and diversity, have been used to summarize the state of an ecosystem. These approaches have numerous shortcomings. We have developed a method for the visualization and quantification of the state of an ecosystem that projects from the original n-dimensional space into a two dimensional representation. Currently, a principal components projection provides the axes to plot the system in a two dimensional space. In studies with several sampling dates, a projection is plotted for each sampling day and then connected to form a three dimensional representation of the changes of the ecosystem over time. The response-volumes or space-time worms generated by this process provide a three dimensional representation of the changes of an ecosystem over time. Various perspectives can be generated until the best viewing point is selected for the particular attribute or question under consideration. The method has proven vital in the examination of microcosm ecosystems dosed with a variety of toxicants and should prove useful in the analysis of FIFRA type microcosms and various field studies. A demonstration of the technique will be presented.
This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.
Key Words: ecosystem effects, space-time worms, response volumes, microcosms

Use of the Mixed Flask Culture (MFC) Microcosm Protocol to Investigate the Effects of a Pulsed Release of Jet-A. R.S. Sandberg and Landis, W.G., Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies; Roze, M.J., Computer Science Department, Western Washington University, Bellingham, WA. A 60-day 1 L Mixed Flask Culture (MFC) microcosm utilizing organisms derived from natural systems was used to assess the potential ecosystem level effects of a simulated release of a complex hydrocarbon mixture from sediments. A spiked layer of Standardized Aquatic Microcosm (SAM) sediment was encapsulated under an overlying layer of coadapted MFC silica sand and detritus. Treatment sediment groups consisting of six microcosm replicates were spiked with 0, 2, 10 and 25 microliters of Jet-A based on the results of preliminary acute 10-day freshwater sediment amphipod bioassays using *Hyalella azteca* as the test species. A slow, pulsed release of the test material from the spiked layer was obtained by stirring vigorously twice weekly throughout the test. Statistically significant effects among both community level physical properties and individual species population dynamics were observed using conventional univariate and multivariate techniques as well as a recently developed non-metric multivariate clustering technique despite the relatively small proportion of Jet-A used in the test.
This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.
Key Words: microcosms, standardized aquatic microcosm, mixed flask culture, sediments.

Evaluation of Community Structure and Community Function After Exposure to the Turbine Fuel Jet-A. S.C. Rodgers and Landis, W.G., Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA. The underlying premises of the Mixed Flask Culture (MFC), an aquatic microcosm design, include 1) that the effects of a perturbation to an aquatic community may be monitored through the measurement of its functional parameters (i.e. pH and productivity/respiration ratio) and 2) these measurements will be similar between different wild-derived communities given the same perturbation. Two MFC experiments were conducted to assess these two premises. The treatment groups in both experiments consisted of 0%, 1%, 5%, and 15% WSF Jet-A with six replicates respectively. The experimental designs reflected both the MFC and the Standard Aquatic Microcosm (SAM); this hybrid design resulted in following a MFC protocol, but incorporated the SAM specified laboratory cultured organisms. Beaker heterogeneity was encouraged in the second experiment by not cross inoculating or reinoculating. The differences between the two experiments was designed to indicate if differently derived communities react similarly to an identical perturbation. Do the microcosms within each treatment group resemble each other functionally throughout the experiment, or is the within group deviation greater than the between group deviation?
This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.
Key Words: mixed flask culture, community function, community structure

Comparison of the Degradation of Water Soluble Components in Jet Fuel Using the Standard Aquatic Microcosm (SAM) and the Mixed Flask Microcosm (MFC). A.J. Markiewicz, Matthews, R.A. and Landis, W.G., Institute of Environmental Toxicology and Chemistry, Western Washington University, Bellingham, WA 98225. The Standard Aquatic Microcosm (SAM), a synthetic assemblage of organisms derived from laboratory cultures, was used in comparison with the Mixed Flask Microcosm (MFC), derived from natural sources, to monitor the degradation rates and biodegradation products of water soluble components in jet fuel and to evaluate whether ecosystem dynamics are similar between the two microcosm systems; independent of species diversity and trophic level complexity. The SAM microcosms were used for analysis of the water soluble fraction of JP-8, and the MFC microcosms were used for the water soluble fraction of Jet-A. Component degradation and by-products were monitored using Purge and Trap / Gas Chromatography. Preliminary results from both microcosms, using regression and multivariate analysis, indicate that all components are degraded simultaneously, but at different rates; component degradation rates oscillate in similar patterns temporally; most WSF components are completely degraded within 10-15 days; and that biodegradation products continue to reappear in a cyclic pattern throughout the experiment. In the SAM microcosms, WSF jet fuel components were rapidly sequestered from the water column and degradative rates were lower. Both microcosms form significantly distinct groups when clustered by degradation rates.
This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.
Key Words: Microcosms, jet fuel, degradation rates.

Abstracts of the 1994 ASTM Symposium on Environmental Toxicology and Risk Assessment, Montreal, Quebec, Canada

Structural and Community Level Comparison of Turbine Fuel Test Results Using the Standardized Aquatic Microcosm (SAM) and the Mixed Flask Culture (MFC) Protocols. Wayne G. Landis, Robin A. Matthews and April J. Markiewicz, Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies; Geoffrey B. Matthews and Michael J. Roze, Computer Science Department, Western Washington University, Bellingham, WA 98225.

The water soluble fraction of the turbine fuels Jet-A, JP-4 and JP-8 have been examined as stressors for two microcosm protocols, the standardized aquatic microcosm (SAM) and the mixed flask culture (MFC). The SAM is a 3 L system inoculated with standard cultures of algae, zooplankton, bacteria, and protozoa. In contrast, the MFC is 1 L and is inoculated with a complex mixture of organisms derived from a natural source. Analysis of the organism counts and physical data were conducted using conventional and newly derived multivariate nonmetric clustering methods, and visualization techniques (space-time worms).

Physical parameters, such as pH and oxygen metabolism, were often not as sensitive as species and bacterial counts. In both the SAM and MFC test systems, species numbers and other variables that determined clusters varied among sampling dates. Compared to the larger yet simpler system, the MFC exhibits more violent dynamics and is more likely to become catastrophically fixated, as in systems dominated by cyanobacteria.

Measurements of the degradation of the various constituents of the water soluble fraction of two jet fuels occurred in both types of microcosms. In these systems apparent shifts in the microbial flora is observable as determined by the release of metabolic products into the media.

Observation of the dynamics using multivariate metric and especially nonmetric clustering reveal similar dynamics at the system level although the structure of the two systems are disparate. In both sets of experiments it appears that an initial divergence is followed by a convergence from some aspects, followed by repeated divergences. The pattern is not as clear in the MFC because of the more rapid shifts in structure. Although both experiments are performed as specified, recovery is not apparent. Recovery is being defined as a return to the original state space and vector, or at least to that of the controls. The inability to clearly separate dosed treatment groups from not dosed treatments in the MFC is more likely due to the rapid divergence of the replicates rather than a recovery or the establishment of an equilibrium.

This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

Community Conditioning as an Alternative to the Stability and Recovery of Ecosystem Hypothesis in Ecological Risk Assessment. Wayne G. Landis and Robin A. Matthews, Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies; Geoffrey B. Matthews, Computer Science Department, Western Washington University, Bellingham, WA 98225.

A common assumption in environmental toxicology is that after the initial stress, ecosystems recover to resemble the control state or reference site. In some instances a new equilibrium state may be established. These assumptions may be based more on our inability to observe an ecosystem with sufficient resolution to detect differences, than reality. Recent findings of complex dynamics in relatively simple microcosms and ecological field studies demonstrate that non-equilibrium systems are the rule rather than the exception.

In a series of microcosm experiments, multivariate analysis and was able to differentiate oscillations that separate the treatments from the reference group, followed by what would normally appear as recovery, followed by another separation into treatment groups distinct from the reference treatment and each other. The initial impact of the toxicant re-sets the dosed communities into different regions of the n-dimensional space where recovery may be an illusion due to the incidental overlap of the oscillation trajectories occurring along a few axes. We now use the construct of space-time worms to visualize the trajectories of the ecosystems through n-dimensional ecosystem space. The dynamics appear to have little regularity and resemble chaotic systems in the lack of repeatability and the importance of initial conditions. However, the systems appear to be bounded, and replicates of a treatment group do appear to follow similar, irregular trajectories. A new vocabulary and an understanding of complex systems (Nicolis and Prigogine, 1989) has been developed that have a direct applicability to community level systems. It is no longer sufficient to say that ecological systems are complex and difficult to understand, an understanding of these systems should be attempted using the fundamentals of complexity.

An outgrowth of this research has been the development of a specific theory, that of Community Conditioning, that generates specific and testable hypotheses regarding ecosystems at the community level. The theory is conservative in that it incorporates many of the characteristics of complex systems. Evolutionary events are incorporated as the "memory" of the ecosystem. This structure along with the exact nature of the toxicant stress must be incorporated into the etiology of the detectable outcome, that according to the community conditioning hypothesis, may be widely separated from the initial stressor event. Community conditioning easily incorporates direct and indirect effects, and actually views indirect effects as a part of the etiology of further outcomes and as adding to the "memory" of the system. Specific hypotheses generated from Community Conditioning include:

1. The complexity and nonlinear dynamics of a biological community may create long latency periods between observable cause and effects.
2. There are patterns in common to communities of different compositions and physical scales. These patterns are more likely to be those at the system level rather than specific interactions among species.
3. The history of an ecosystem is essential in determining the etiology of an effect due to a toxicant or other stressor.

Community conditioning is an alternative hypothesis and model for the interpretation of toxicant impacts and for assessing risk at the community level. It is an alternative to the stability and recovery model of ecological response to stressors.

This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

Nicolis, G. and I. Prigogine (1989) *Exploring Complexity: An Introduction.* W. H. Freeman and Company, New York,.

Artificial Intelligence Based Data Analysis and Visualization Tools for Ecological Risk Assessment . Geoffrey B. Matthews, Computer Science Department; Robin A. Matthews and Wayne G. Landis, Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA 98225.

Data analysis in ecotoxicology is hampered by the lack of sophisticated visualization tools. The data are typically complex, multidimensional and time-dependent. We have found that sophisticated visualization techniques can speed data interpretation by factors of two to ten. Many other disciplines, such as geography and meteorology, have devoted considerable energy to the development of computer visualization tools to enhance understanding of complex data (e.g., GIS, Geographic Information Systems). Computer workstations are falling in price and increasing in power on an almost daily basis. Desktop, three dimensional, interactive, real-time data visualization is now a reality. It is time that ecotoxicology develop a suite of approaches and tools that can become a standard part of laboratory and field multivariate data investigation.

In this paper, we present an overview of some of our visualization tools, and their applicability to ecotoxicological data analysis and ecological risk assessment. These tools are now ported to a standard 486 computer running the NEXTSTEP operating system. A unique feature of our tools is that they are integrated with our artificial intelligence (AI) software, to aid the investigator in understanding as well as visualization. Currently the AI and visualization tools are being integrated as an executive software program called MuSCLE (Multivariate Software for Community Level Ecology).

We survey the approaches, present our results from analysis of real field and laboratory studies, and demonstrate the utility of the software. Studies of eutrophication, enrichment of stream ecosystem, laboratory microcosms (Standardized Aquatic Microcosms and Mixed Flask Culture) and biomarker data from field plot studies will be presented. In each case, features of the data were revealed that were not visible otherwise, and in some cases, have led to new interpretations of the impacts of toxicants upon ecological systems.

This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

Annual Meeting of the Pacific Northwest Chapter of the Society for Environmental Toxicology and Chemistry, University of Victoria, Victoria, B.C., May 1994

THE RECOVERY MYTH AND AN ALTERNATIVE-COMMUNITY CONDITIONING. W. G. Landis, R. A. Matthews, Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies; G. B. Matthews, Computer Science Department, Western Washington University, Bellingham, WA 98225.

A common assumption in environmental toxicology is that after the initial stress, ecosystems recover to resemble the control state or reference site. These assumptions may be based more on outmoded theory than reality. Recent findings of complex dynamics in relatively simple microcosms, chaotic dynamics in ecological field studies, and techniques of examining complex datasets demonstrate that non-equilibrium systems are the rule.

The use of nonmetric clustering in the analysis of ecological datasets has led us to formulate a non-equilibrium theory, the community conditioning hypothesis. The community conditioning hypothesis states that ecological communities preserve information about every event in their etiology. In our studies of standardized aquatic microcosms (SAMs), for example, we observed distinct community changes in response to stress that would appear and disappear over a two-month period (Landis et al., 1993a; Landis et al., 1993b). Even after the dosed systems had "recovered" to a state indistinguishable from the reference systems, a stress effect reappeared. A purely stochastic system could not exhibit this effect, since information is erased over time and two systems with identical distributions will remain identically distributed. A chaotic system could exhibit this effect, but we do not believe these microcosms are inherently chaotic, since similar systems tend to follow similar evolutions, without the divergences characteristic of nonlinear systems. Instead, we advanced the hypothesis that an unobserved feature of the community carried information about the stressor throughout the history of the system. In the case of the SAMs, we hypothesize detrital conditioning as the mechanism by which information is preserved. The preservation of the information can be contained in a variety of structural components of the ecological system, including genetics, competitive interactions, migration dynamics, community structure or age structure of a population. Examples of such conditioning can be found in the mitochondrial sequences of human populations and the affinity of 2,3,7,8 dioxin for the vertebrate Arh receptor.

An outgrowth of this research has been the development of a specific theory, that of Community Conditioning, that generates specific and testable hypotheses regarding ecosystems at the community level. Specific hypotheses generated from Community Conditioning include:

1. Biological communities may have long latency periods between observable cause and effects.
2. Communities may have patterns in common despite differences in compositions and physical scales. These patterns are more likely to be those at the system level rather than population level.
3. The history of an ecosystem is essential in determining the etiology of an effect.

Community conditioning is an alternative hypothesis and model for the interpretation of toxicant impacts and for assessing risk at the community level. It is a clear and discrete alternative to the stability and recovery model of ecological response to stressors.

This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

Landis, W.G., R.A. Matthews, A.J. Markiewicz, N.A. Shough and G.B. Matthews. (1993a). Multivariate Analyses of the Impacts of the Turbine Fuel Jet-A Using a Microcosm Toxicity Test. *J. Environ. Sci.* Vol 2:113-130.

Landis, W.G., R.A. Matthews, A.J. Markiewicz and G.B. Matthews. (1993b) Multivariate Analysis of the Impacts of the Turbine Fuel JP-4 in a Microcosm Toxicity Test with Implications for the Evaluation of Ecosystem Dynamics and Risk Assessment. *Ecotoxicology* 2:271-300.

Abstracts Submitted for SETAC 94, Denver, Colorado

The Inherent Limitations of Population Modeling in Environmental Risk Assessment and an Alternative: Community Conditioning. W.G. Landis and R.A. Matthews, Institute of Environmental Toxicology and Chemistry; G.B. Matthews, Computer Science Department, Western Washington University, Bellingham, WA. As recently pointed out by Oreskes et al (*Science* 1994, 263: 641-646) explanation based models can not be validated or verified, only confirmed. In addition to this challenge, we have observed phenomena in multispecies toxicity tests that is not adequately described by stochastic or deterministic models. After microcosms had "recovered" to a state indistinguishable from the reference systems, a stress effect reappeared indicating that information about the prior stressor remained. Trajectories are sensitive to initial conditions. A purely stochastic model could not produce this effect, since information is erased over time and two systems with identical distributions will remain identically distributed. A chaotic model could exhibit this effect, but we do not believe these microcosms are inherently chaotic, since similar systems tend

to follow similar evolutions, without the divergences characteristic of chaotic nonlinear systems. Instead, we advance the hypothesis that often unobserved features of the community carry information about stressor events throughout the history of the system. Thus ecological systems are irreversible and that historical components are critical. These assumptions form the core of the Community Conditioning Hypothesis. As an alternative, similarity based models, like conceptual clustering, may have an important role in the prediction of risks in ecological systems. (USAFOSR Grant No. AFOSR-91-0291 DEF)

Application of the Community Conditioning Hypothesis to the Design of Multispecies Toxicity Tests. R.A. Matthews, W.G. Landis, Institute of Environmental Toxicology and Chemistry, and G.B. Matthews, Computer Science Department, Western Washington University, Bellingham, WA. Multispecies toxicity tests are used to provide a more realistic assessment of a toxin's environmental risk than single-species toxicity testing. Multispecies test typically measure changes in physical-chemical parameters as well as in the number, biomass, or growth rate of selected macroinvertebrate, algal, and fish taxa. Ecologists, however, point out that communities are complex systems and that observable populations are strongly influenced by unobservable phenomena, such as microbial interactions or detrital conditioning. We have observed population responses in SAM microcosms that seem to be unexplainable without postulating such hidden factors: communities that appear to have recovered from toxic stress after one month nevertheless show significant responses after two months. Based on these observations we have proposed the Community Conditioning Hypotheses: ecological communities retain information about their history indefinitely. The following factors, therefore, are important additions to the design of multispecies toxicity tests: the test duration should be long enough to allow the development of a detrital food base and to expose latent, dose-related population responses, and "key" microbial processes (including, but not limited to, toxin degradation rates) should be measured. Further, no assumption of "recovery" should be made for any community simply on the failure to detect a significant difference between treatment groups, for unobservable differences can have observable consequences.

The Stability Myth and the Dynamics and Patterns of Xenobiotic Impacts to Ecological Systems W.G. Landis, R. A. Matthews, Institute of Environmental Toxicology and Chemistry, M. A. Roze. and G. B. Matthews, Computer Science Department, Western Washington University, Bellingham, WA. A common assumption in environmental toxicology is that after the initial stress, ecosystems recover to resemble the control state or reference site. Recent research in the organization of community structure, chaotic dynamics in ecological field studies, and techniques of examining the dynamics of complex datasets demonstrate that non-equilibrium systems are the rule. In our microcosm studies, we have observed distinct community changes in response to stress that would appear and disappear over a two-month period. Even after the dosed systems had "recovered" to a state indistinguishable from the reference systems, a stress effect reappeared. Neither stochastic or chaotic dynamics can adequately describe the observed phenomena. We advance the hypothesis of community conditioning, that an unobserved feature of the community carries information about the stressor throughout the history of the system. The preservation of the information can be contained in a variety of structural components, including genetics, competitive interactions, migration dynamics, age structure of a population or community structure. Community conditioning is an testable hypothesis and model for the interpretation of toxicant impacts and for assessing risk at the community level. It is a clear and discrete alternative to the stability and recovery model of ecological response to stressors. (USAFOSR Grant No. AFOSR-91-0291 DEF)

An Multivariate Artificial Intelligence Approach to the Evaluation of Biomarkers Under Field Conditions II. W.G. Landis, Institute of Environmental Toxicology and Chemistry, M. A. Roze. and G. B. Matthews, Computer Science Department, Western Washington University, Bellingham, WA., S. Dominguez, U. S. Environmental Protection Agency, Corvallis OR, A. Fairbrother, ecological planning and toxicology Inc., Corvallis OR. The biomarker data were derived from field experiments using gray-tailed voles placed in 0.2-ha field enclosures and dosed with azinphos methyl. Molecular markers included brain cholinesterase activity, blood chemistry, enzymatic and cell type markers. Data were analyzed using nonmetric clustering and association analysis (NMCAA), an artificial intelligence technique. NMCAA confirmed the ANOVA results in that brain cholinesterase activity was an important variable in clustering on treatment group. However, NMCAA found that neutrophils and basophils were also important variables. The alteration of the ratio of leukocyte types has been previously reported in laboratory tests with azinphos methyl. The quality measurements of the clustering suggested that additional patterns are present. Using quadrants within the field experiment as a treatment, a statistically significant relationship was again found. The variables determined to be important were brain cholinesterase inhibition, pregnancy and basophils. Controlling for quadrant effects, again a significant association between dose and clusters was found within the quadrants even though the total sample size within a quadrant was significantly reduced. The biomarkers and NMCAA detected at least two important patterns in the field experiment, dose and location. (USAFOSR Grant No. AFOSR-91-0291 DEF).

APPENDIX A

Publications 1991-1994
and Submitted Manuscripts

INSTITUTE OF ENVIRONMENTAL TOXICOLOGY AND CHEMISTRY
Huxley College of Environmental Studies, Western Washington University
Bellingham, WA 98225-9180

10 August 1993

Dr. Jeffrey M. Giddings
Editor, SETAC News
Springborn Laboratories, Inc.
790 Main Street
Wareham, MA 02571

Dear Editor:

We have been following with great interest the discussion regarding the decision by USEPA to reduce the number of field tests and aquatic FIFRA microcosms required for the registration of pesticides (Fisher, L., Oct 29, 1992 USEPA). In the July 1993 issue of the SETAC News Dr. Frieda Taub asks the question, **"Can we make a sufficient case to demand that EPA re-evaluate their stand and provide funds for an initiative to develop appropriate ecosystem-level testing for regulatory purposes?"**

The answer to the question is yes, but not as a request to return to the status quo of pre-1992. Several conceptual breakthroughs have been made in the last few years that make a return to ecosystem level testing a realistic part of the regulatory process. New methods of data analysis and visualization make it practical to evaluate the systems as a entity, freeing the field from reliance on analysis of variance and derived methods . In addition, a new vocabulary and an understanding of complex systems (Nicolis and Prigogine, 1989) has been developed that have a direct applicability to community level systems. It is no longer sufficient to say that ecological systems are complex and difficult to understand, an understanding of these systems should be attempted using fundamentals of complexity.

Multispecies systems whether field studies, FIFRA mesocosms or Standardized Aquatic Microcosms all have properties fundamentally different from single species toxicity tests. These properties are not all unique to living systems but fall into the realm of complexity. Complex systems are nonlinear in nature, may produce chaotic dynamics, and all incorporate changes that reflect the consequences of historical events, and are irreversible. The irreversible property of complex systems are often manifested in a cascade of direct and indirect effects. As the direct effects of the toxicant manifest themselves, that information is imparted to the system through a sequence of indirect effects. These indirect effects are initiated as soon as the direct effects become manifest. The initial populations of the affected species are altered, predator-prey relationships change, processing and recycling of detritus transformed, as well as selection for certain resistant genotypes that alter the population genetics of the surviving populations. As the toxicant degrades it is the indirect effects that carry the consequences of that toxicant impact for an undetermined time. Indirect effects are both immediate and long-term. Depending upon the exact nature of these changes, the resultant system may diverge significantly from the desired state or be more or less sensitive to additional stressors.

Recovery as a return to the original condition or to that of a so called reference site may simply not exist as a property of these systems. Indeed, evidence for such properties has often proven underwhelming (Connell and Sousa 1983). Perhaps a more workable definition of recovery may be the inability to distinguish the impacts due to environmental perturbations from those due to the historical pollutant stress. Although the impacts of the stressor event remain as part of the overall information content of the system, it is overwritten to some degree by other environmental factors.

Our current research with microcosms (Landis et al in press), streams and lakes, and even molecular and physiological markers has demonstrated the power of data analysis methods derived from artificial

intelligence research approaches coupled with and the power of complexity theory in the formulation of specific and general hypotheses. Modeling using classical physical models may be inappropriate for complex systems, and certainly can not take the place of the experimental confirmation of theory. Indirect effects are immediate and are the most persistent part of the impact of a stressor. These indirect effects actually seem to form a "memory" of that event within the system. Finally, as with most complex systems, ecosystems are irreversible, and definitions such as stability and recovery should perhaps be reconsidered as to their utility and even appropriateness.

It is inappropriate to eliminate testing at the field or multispecies level. All that remains is inference from dissimilar systems and models without confirmation. The inability to apply the existing paradigm to field research does not mean a retreat from the ultimate object of study and protection should occur. Instead, incorporation of the ideas presented here along with many other concurrent developments should spur a return to multispecies systems and field research. If the paradigm shifts, so be it.

We'll all be at SETAC Houston and certainly will enjoy a follow on discussion.

References:

Connell, J. H. and Sousa, W. P. (1983) On the evidence needed to judge ecological stability or persistence. *Amer. Nat.* 121:789-824.

Landis, W. G., R. A. Matthews, A.J. Markiewicz, N. A. Shough and G. B. Matthews. *In press.* Multivariate Analyses of the Impacts of the Turbine Fuel Jet-A Using a Microcosm Toxicity Test. *J. Environ. Sci.* Vol 2:2.

Landis, W. G., R. A. Matthews, A. J. Markiewicz and G. B. Matthews. *In press.* Multivariate Analysis of the Impacts of the Turbine Fuel JP-4 in a Microcosm Toxicity Test with Implications for the Evaluation of Ecosystem Dynamics and Risk Assessment. *Ecotoxicology.*

Nicolis, G. and Prigogine, I. (1989) Exploring Complexity: An Introduction. W. H. Freeman and Co., New York.

Sincerely,

DR. WAYNE G. LANDIS, Director
Institute of Environmental Toxicology
and Chemistry
Internet: landis@henson.cc.wwu.edu

DR. ROBIN A. MATTHEWS
Institute of Environmental Toxicology
and Chemistry
Internet: robin@duckweed.ietc.wwu.edu

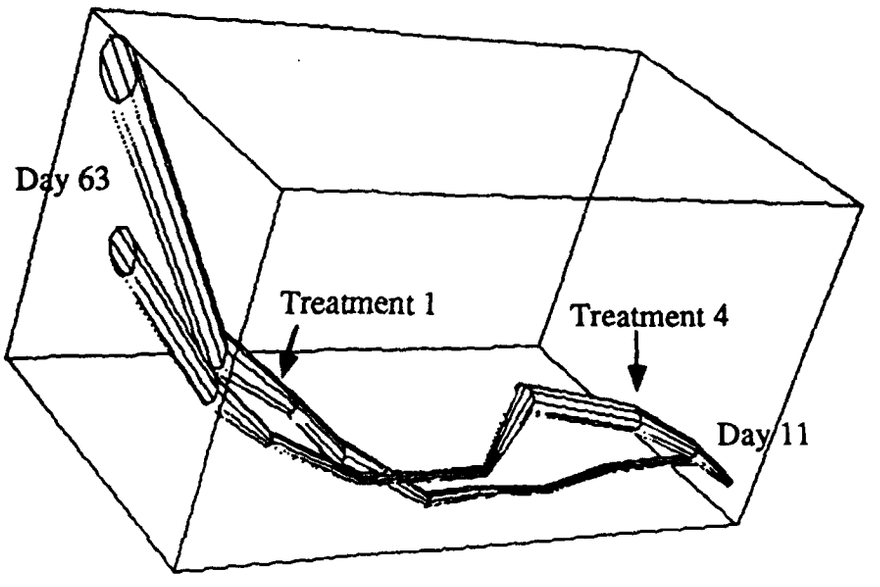
DR. GEOFFREY B. MATTHEWS
Department of Computer Science
Internet: matthews@fortress.cs.wwu.edu

Note to Jeff Giddings: I hope this is not too technical or too long for publication. I felt the references are warranted just for background to some of the topics. I'll fax a copy to Frieda Taub for her comments.

Nonmetric Clustering, Microcosms, Spacetime Worms, Complexity and their Application to Ecological Risk Assessment

Papers and Preprints 1990-1994

Institute of Environmental Toxicology and Chemistry
Huxley College of Environmental Studies, Western Washington University
Bellingham, WA 98225-9180



Response Area (Wormspace) for the Jet-A SAM Experiment

Nonmetric Clustering, Microcosms, Spacetime Worms, Complexity and their Application to Ecological Risk Assessment

Papers and Preprints 1990-1994

Huxley College of Environmental Studies
Western Washington University
Bellingham, WA 98225-9180
Phone: 206-650-6136
Fax: 206-650-7284
Internet: landis@henson.cc.wwu.edu

Dr. Wayne G. Landis
Director

June 1994

CONTENTS

•**Nonmetric Conceptual Clustering in Ecology and Ecotoxicology**--G. Matthews, R. Matthews, and W. Landis, In Press, *AI Applications.*

•**Application of Multivariate Techniques to Endpoint Determination, Selection and Evaluation in Ecological Risk Assessment**--W.G. Landis, G.B. Matthews, R.A. Matthews and A. Sergeant, In Press, *Environ. Tox. & Chem.*

•**Comparison of the Community Dynamics of the Standardized Aquatic Microcosm and Mixed Flask Culture Microcosm Toxicity Tests with Suggestions for Design Criteria in Multispecies Toxicity Testing**--W.G. Landis, A.J. Markiewicz, R.A. Matthews, M.J. Roze and G.B. Matthews, Submitted, Ecological Risk Assessment of Pesticides: Enhancing the Process (SETAC Special Publication).

A Model for Describing Community Change[1]

Geoffrey Matthews
Computer Science Department

Robin Matthews
Huxley College of Environmental Studies
Western Washington University

U.S. EPA Conference on Fate and Effects of Pesticides in the Environment, December, 1990, Corvallis, OR

ABSTRACT

Biological monitoring and multispecies toxicity tests generate complex, multivariate data sets. The primary tools found useful in studies of multivariate data have been ordination and classification techniques based on a view of the data matrix as a collection of points in a highly-dimensioned feature space. This view usually requires making unsupported assumptions about the data (Gaussian distributions, equal variances, etc.) Where these assumptions are not met, it is often necessary to transform the raw data by taking logarithms, normalizing the variances, or eliminating outliers. We have developed a technique (Clustering and Association Analysis) that measures the strength of associations between clusters and treatment groups (or samples grouped by location, date, etc.). In our technique the data are first clustered independently of their treatment group. We advocate the use of nonmetric clustering for this step because it is insensitive to changes in scale and can filter out many effects due to outliers and differences in variance between parameters. After the clusters are generated, the degree of match between the clusters and the treatment is calculated. If the data are strongly influenced by the treatment, the clusters in the data will have a strong association with the treatment. On the other hand, if the treatment or location has no effect, the clusters will be random with respect to treatment. The strength of this association can be used to determine a significance level for the effect. We present the results of this technique on data from a standardized aquatic microcosm (SAM) test.

INTRODUCTION

Biological monitoring and multispecies toxicity tests (microcosm and mesocosm) continue to grow in importance. They address the problems of community change, and the analytical tools used to study them must be constructed in this light. Measurements on dozens to hundreds of species and abiotic parameters result in complex, multivariate data sets. The peculiarities of environmental monitoring result in problems for the analysis of this data, as well. Many species are absent, resulting in many zeroes in the data matrix. Rare species and common species may each indicate effects, although their variances are quite different. Counts may be in individuals, clusters, or colonies. Observations are quite often simply "missing" or incomplete, due to hazards of field work.

In this paper we advocate a methodology for analyzing such data sets with the express goal of simplifying the data. We want to reduce the data to its important aspects. We do this in two ways. First, the samples, which usually run into the hundreds, are reduced into a few fundamental clusters. Second, the measured parameters, both biotic and abiotic, will be reduced to a few important ones. The important ones are simply those which have the strongest association with the sample clusters. We present the essentials of our technique in the context of discussing the analysis of data from a standardized aquatic microcosm experiment.

A Standardized Aquatic Microcosm Study

The standardized aquatic microcosm test we use here for illustration involved the testing of a toxin, and also the possible mitigating effects of a bacterium which degraded the toxin. The toxin was CR, a riot control chemical, and the bacterium is known as CR-1. Questions about the SAM test itself should be directed to Wayne Landis, Institute for Environmental Toxicology and Chemistry, Western Washington University.

[1]We wish to thank Wayne Landis, Institute of Environmental Toxicology and Chemistry, Huxley College, Western Washington University, for his contributions to our project and for providing the SAM study data.

The experiment was set up with four treatment groups, and two flasks in each group. Flasks 1 and 2 were the control group, flasks 3 and 4 had the toxin added, flasks 5 and 6 had the bacterium, but not the toxin, added, and flasks 7 and 8 had the toxin and the bacterium. Typical biotic responses to the test are shown in Figure 1, and abiotic parameters in Figure 2. As can be seen by looking at the response of *Daphnia* in Figure 1, the degradation products of this toxin were also toxic. In flasks 3, 4, 7 and 8, the *Daphnia* die out after administration of the toxin, while they show very healthy growth in flasks 1, 2, 5 and 6, where no toxin was administered. A secondary effect on the algae can be seen in the response of *Ankistrodesmus* in Figure 1. The absence of the predator, *Daphnia*, in the toxic groups allows *Ankistrodesmus* to enjoy healthy growth.

Examination of the data by eye thus reveals that although there were four treatment groups, there were really only two responses to the four treatments. We wish to find an analytical tool which will confirm this, or, indeed, reveal it in cases where it is not obvious to the eye, and also give us some indication of which species are significantly associated with this effect. In larger tests, and in field studies, the number of samples and the number of species may be orders of magnitude larger, and the overall effect may be difficult to discern.

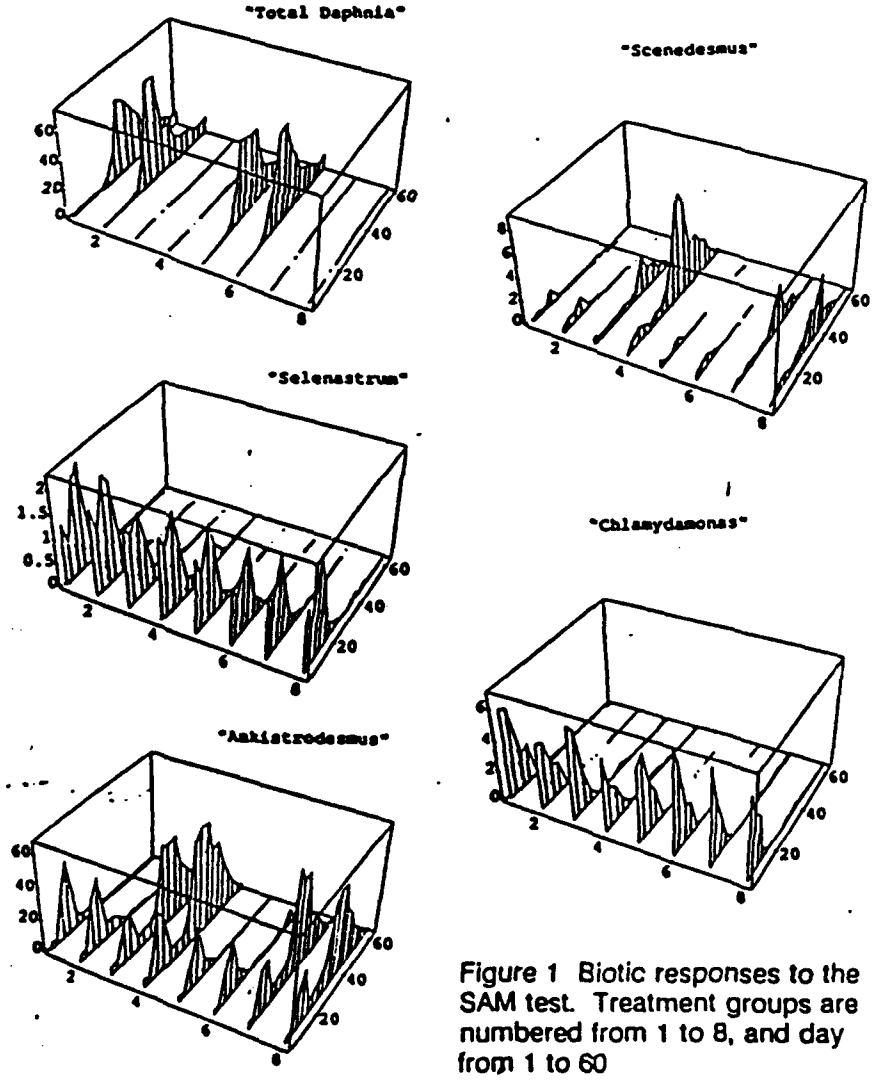


Figure 1 Biotic responses to the SAM test. Treatment groups are numbered from 1 to 8, and day from 1 to 60

Standard Approaches To Multivariate Analysis

There are many approaches to analytically expressing the observed differences between treatment groups or site locations. Some of these approaches are primarily graphical in nature, such as principal components and detrended correspondence analysis, which are designed to reduce the multivariate data to two dimensions which can be inspected and interpreted directly. These techniques, however, still rely on human judgement to determine the strength and nature of possible effects.

Another common approach to multivariate data is to try to reduce a sample, with its associated measures on many species, to a single number which combines all these numbers into one. The Shannon-Weaver diversity index is an example. One problem with this approach is simply that it often does not work. In our example SAM study, the diversity indices are plotted in Figure 3, and there does not appear to be any strong indication of two responses to the four treatment groups.

Another approach to understanding multivariate data is to view each sample, with its associated measurements on many parameters (species, temperature, pH, etc.), as a point in n-dimensional space, where n is the number of parameters. This will permit summary statistics about groups, which are collections of sample points, in terms of metric properties about a collection of points in n-space. This is the background to a wide

variety of approaches, including multivariate analysis of variance (MANOVA) and an approach based on similarity measures (Smith et al., 1990). These approaches show a great deal of promise, but their reliance on the n-dimensional metric approach to multivariate data leaves them all subject to certain problems. First, there is the choice of metric function itself: Euclidean distance, squared Euclidean distance, cosine of vectors distance, Mahalanobis distance, and many others all have various features to recommend them, but the choice of a particular one for a particular problem remains a difficult decision. Second, there is the sensitivity of many of these metrics to scale. If we change one parameter, for instance, from millimeters to centimeters, we may well change important distances in n-space. If we normalize all measures beforehand, for instance by requiring unit variance, we face the problem of justifying this distortion of the data. For example, it may well be that a particular species has very small variance over all groups. We are then faced with the decision: do we "normalize" this species and magnify its variance to be in line with the other species, or do we make the decision to remove this species from the data set before analysis? Either decision has its problems. Third, there is the problem of incommensurable parameters. Most of the n-dimensional metrics require combining parameters in some fashion, for example, by summing the squares. If the data set is very mixed, however, what is the justification for combining, say, temperature and pH? How can we meaningfully sum the squares of counts for algae, fish, and clams? Worse, how can we combine biotic and abiotic measures? In any event, what do such n-dimensional distances mean?

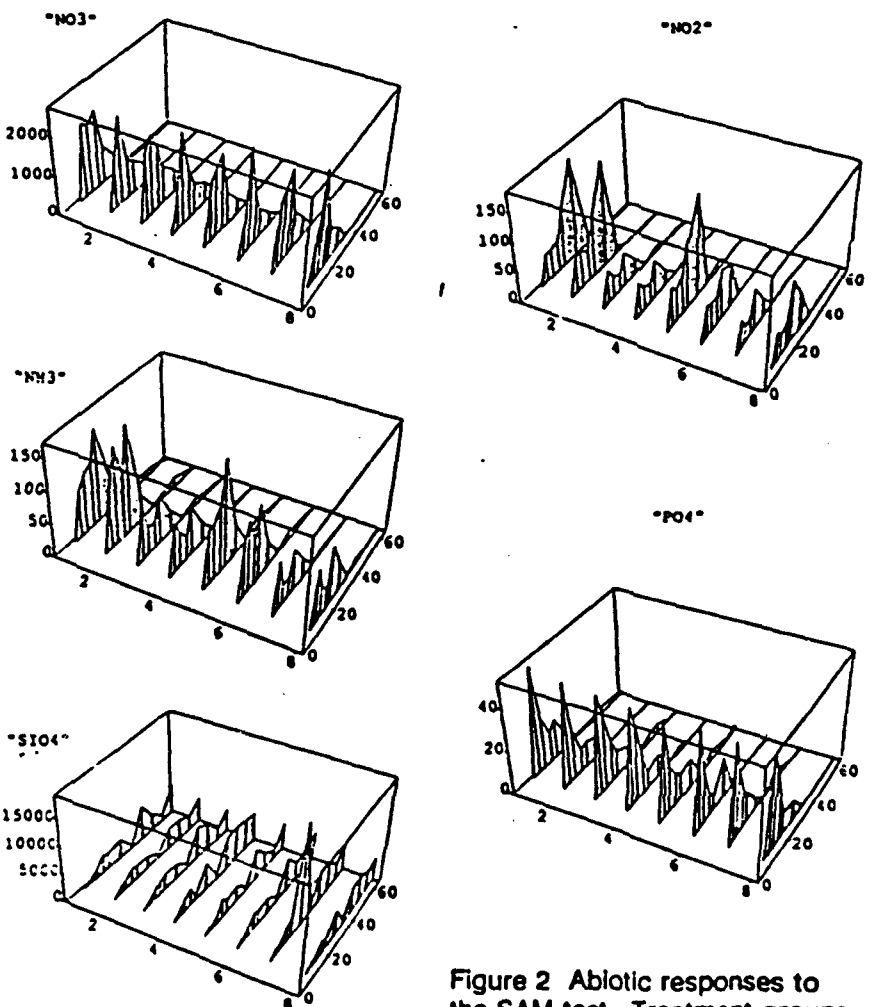


Figure 2 Abiotic responses to the SAM test. Treatment groups are numbered from 1 to 8, and day from 1 to 60

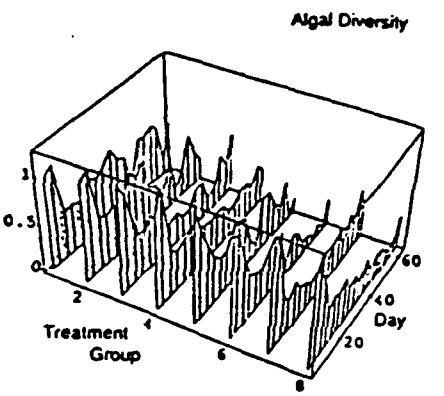


Figure 3 The difficulty with single-number indices, such as algal diversity, as characterizations of community structure is illustrated here for the SAM test

In our work we have strived to avoid the twin pitfalls of oversimplification (as in diversity indices) and a complex approach involving n-dimensional metrics which are difficult to interpret.

Nonmetric Clustering

Our approach is based, first, on nonmetric clustering (Matthews and Hearne, 1991), which we will outline briefly here. Clustering is, first, a technique of pattern recognition. The idea is that a data set of many points may contain patterns or clusters, i.e. a few sets of very similar points. Describing a data set as 100 samples from each of three

clusters is simpler, and more accurate, than describing the same data set as simply 300 points. Therefore, the recognition of these patterns in complex data is paramount to understanding it on a deeper level.

Traditional clustering algorithms, unfortunately, rely on distance measures or metrics in n-dimensional space, just like the approaches discussed above. A set of points is divided into several clusters based on the criterion that the average within-cluster distance should be smaller than the average between-cluster distance. In other words, two points are "similar", or "belong to the same pattern" if their n-dimensional distance is "small". The differences between the various algorithms for clustering, agglomerative or divisive, hierarchical or partitioning, are mainly in how these clusters are found. But in each case, the criterion for clustering validity still relies on an n-dimensional distance or similarity function. For the reasons advanced in the previous section, nonmetric clustering was developed as a pattern recognition technique which avoids reliance on n-dimensional metrics.

The primary distinction of nonmetric clustering is its definition of clustering validity: a clustering of data points is good if the data and the clusters are strongly associated. In other words, if you know which cluster a data point belongs to, then you have a good idea of what kinds of data values it will have. Suppose, for instance, that the SAM data (Figures 1 and 2) were divided into two clusters, where samples from flasks 1, 2, 5 and 6 were in cluster "A" and samples from flasks 3, 4, 7 and 8 were in cluster "B". Then you would know that if a sample were from cluster "A" it would, by about day 35, have large numbers of *Daphnia* and small numbers of *Ankistrodesmus*, and vice versa if it were from "B". There may be some parameters about which you know little, for example *Chlamydamonas*, but the important thing about a good clustering is that, at least for some parameters, it gives you a good idea about the values for the points in the clusters.

We have implemented nonmetric clustering in a computer program called RIFFLE (Matthews and Hearne, 1991). This HP LaserJet II D (25 IN ONE 103)HLIIDADD.PRSetween clusters and parameter values of the data points. The strongest association between clusters and parameters, for the largest number of parameters, gives the best clustering. We have used this clustering program on a wide range of data sets, and have found it to be consistently superior to traditional clustering algorithms (Matthews, Matthews and Ehinger 1991; Matthews, Matthews and Landis, 1990; Matthews, Matthews and Hachmoller, 1990; Matthews, 1988). In the case of the SAM data, a nonmetric clustering on day 35 showed that, indeed *Daphnia* and *Ankistrodesmus* were strongly associated with the best clustering. Thus, nonmetric clustering achieves both halves of the data reduction task: the samples are reduced to a few clusters, and the parameters are reduced to those few which are best associated with the clusters. In the SAM case, and in many of our other tests, the parameters selected by nonmetric clustering as the most significant are in concert with the ones a human expert would select.

Clustering and Association Analysis

Clustering is only the first step in the analysis of monitoring and multispecies toxicity test data. The clustering is done independently of the treatment groups (or locations, etc.). Clustering thus identifies patterns in the data without judging whether these patterns are due to, or even associated with, the treatment groups. The next step is to analyze the association between the clusters and the groups. A strong association between groups and clusters indicates a significant effect associated with the treatment or location.

In our SAM data, nonmetric clustering on day 35 divided the samples into two clusters, one consisting of all samples from flasks 1, 2, 5 and 6, and a second cluster consisting of all samples from flasks 3, 4, 7 and 8. In other words, a perfect division of the samples into clusters "with" and "without" the toxin. Since the clustering was done "blind" with respect to the actual treatment groups, this is a striking result. Under the null hypothesis, i.e. that the treatment had no effect on the clustering, such a match between groups and clusters is far less than 1% probable, leading us to reject the null hypothesis at the 99% confidence level.

To make sure our analysis was not biased in favor of two clusters, we clustered the samples on each sampling date into two, three, four, and five clusters. If the four treatment groups had led to, say, four different responses, then the association between the four treatment groups and four clusters would be higher than the association between the four treatment groups and two clusters. As it turned out (Figure 4) association analysis shows that the strongest association was with only two clusters.

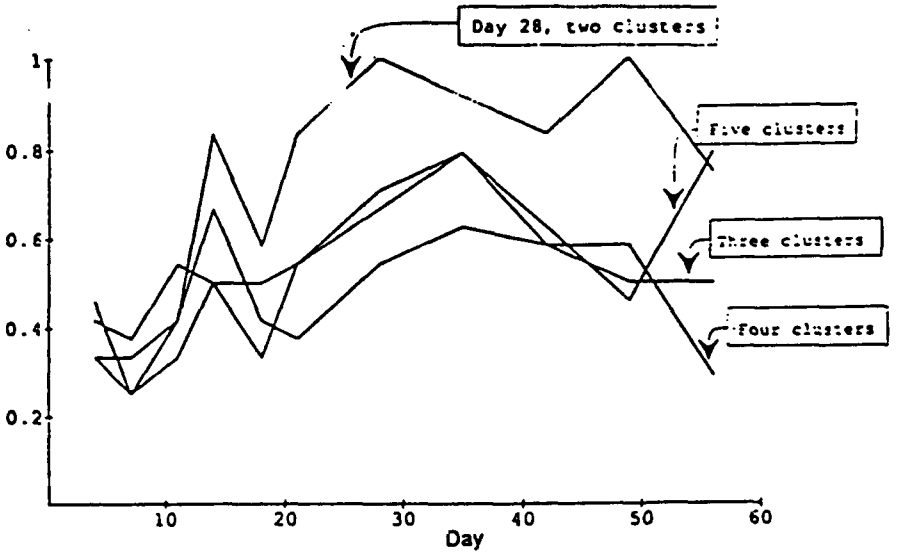


Figure 4 Significance of match between blind clustering by the Riffle algorithm and actual treatment groups. Optimal clustering is achieved using two clusters on day 28. Failure to find a significant association for more than two clusters supports the hypothesis of toxicity for degradation products.

CONCLUSIONS

Clustering and association analysis is based on the answers to the following questions:

1. Are there patterns in the data?
2. Are these patterns associated with the treatment groups?

The answer to the first question tells us whether there is anything "happening" in the data at all. The answer to the second question tells us whether the treatment groups are associated with this effect. One of the benefits of this division into two separate questions is that nonmetric clustering can be used in the pattern recognition phase and so n-dimensional metrics need not be used.

Finally, we would like to point out that traditional significance testing is implictly post hoc. It attempts to determine only whether or not a difference exists between two given populations, the treatment and control groups. If there are, in fact, patterns in the data, traditional testing will not reveal them unless they are associated with the given treatment groups. Our approach, however, looks for patterns in the data independently of the known treatment groups. This pattern analysis of the data can sometimes identify effects that the researcher did not know about; it can give him or her "surprises" and reveal new directions in research. In other words, traditional tests can tell you "yes" or "no" regarding the questions you ask. They cannot tell you "yes, but ...".

LITERATURE CITED

Matthews, Geoffrey. 1988. Clustering Heterogeneous Ecological Data. Annual Conference, International Society for Ecological Modelling. Davis, California.

Matthews, Geoffrey and James Hearne. 1991. Clustering without a metric. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(2).

Matthews, Geoffrey, James Hearne and Peter Sugarman. 1987. Conceptual Clustering in the Analysis of Environmental Data Sets. NOAA Conference on Artificial Intelligence Research in Environmental Science. Boulder, Colorado.

Matthews, Geoffrey, Robin Matthews and Wayne Landis. 1990. Applications of nonmetric clustering (NMC) to pattern analysis in aquatic toxicology. Society of Environmental Toxicology and Chemistry, 11th Annual Meeting. Arlington, Virginia.

Matthews, Robin, Geoffrey Matthews and Barnard Hachmoller.1990. Ordination of Benthic Macroinvertebrates Along a Longitudinal Stream Gradient, Annual Conference, North American Benthological Society. Blacksburg, Virginia.

Matthews, Robin, Geoffrey Matthews and William Ehinger. 1991. Classification and Ordination of Limnological Data. Ecological Modelling (In Press).

Smith, Eric P., Kurt W. Pontasch and John Cairns Jr. 1990. Community similarity and the analysis of multispecies environmental data: a unified statistical approach. Water Res. 24(4): 507-514.

Questions for Geoffrey Matthews:

Q. This model, is it user friendly, or do you have to have a computer degree to use it? And, is it published?

A. It is in press, to answer your last question. It is not production quality. I wouldn't be real proud of it if I released it right now. I'm not trying to hide it, I just don't want to be embarrassed when people look at my software. We are working on putting all the bells and whistles on it.

Q. Is it PC or mainframe based?

A. The search, the clustering, takes a long time. Its search takes a long time. When doing the clustering, it looks through lots and lots of clusters. You can run it on a PC, its written in "C".

Q. How can I get a copy?

A. I'll give you my card. It'll take a long time on a PC. Probably do OK on a 386.

Q. In your method, I like the analogy of putting a nozzle on a fire hose, and if you put a nozzle on a fire hose, that nozzle is metric, even though you say it's non-metric. Can your method be summarized into throttling the flow into something that can explain more things, better things?

A. There are two data reductions that are important, and they are both present here. One is the reduction of the number of variates, the number of species, from 100 down to 5, or 2; 3 in the present case. That's one of the funnels. The other is, instead of 100's of points, 100's of samples, you have 2. Even though you have 100 samples, 50 from here and 50 from here. The important difference is between this bunch and this bunch, so you go from 100 to 2. You are reducing the number of points; you are reducing the number of variates.

Q. But what I meant is, do you have a method of [?] saying, "Hah! Here it is. I didn't know it!"

A. No, then you have to go to the ecologist. I'm the mathematician. I don't do any "Hah!" stuff. (laughter) All my stuff is boring. The exciting part is Wayne's and Robin's. But it does in fact lead to those things.

Q. But does it tell you whether you are right or not?

A. It does more than that. It will tell you surprising things. It will give you "A-hah!'s" You have to be a scientist to recognize them. Like the time nitrate came out. It said "*Daphnia simodesmus*", and then "nitrate". I said, "Robin, why is nitrate here?", and she said "Hmm! I'm not sure. Maybe its nutrient limiting, or something like that." So all of a sudden she was thinking about something she had not thought about before. This will tell you things that you may not have seen before.

Q. What if you change the scales?

A. You will get exactly the same results, if you change the scale on any or all of the variates. It doesn't depend on that.

T-PAM/13/2//40947

Clustering Without a Metric

Geoffrey Matthews
James Hearne

Reprinted from
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE
Vol. 13, No. 2, February 1991

Clustering Without a Metric

Geoffrey Matthews and James Hearne

Abstract—**We describe a methodology for clustering data in which a distance metric or similarity function is not used. Instead, clusterings are optimized based on their intended function: the accurate prediction of properties of the data. The resulting clustering methodology is applicable, without further ad hoc assumptions or transformations of the data 1) when features are heterogeneous (both discrete and continuous) and not combinable, 2) where some data points have missing feature values, and 3) where some features are irrelevant, i.e., have large variance but little correlation with other features. Further, it provides an integral measure of the quality of the resulting clustering. We have implemented a clustering program, RIFFLE, in line with this approach, and experiments with synthetic and real data show that the clustering is, in many respects, superior to traditional methods.**

Index Terms—**Clustering, cluster validity, multivariate data, proximity indexes, unsupervised learning.**

I. Introduction

THE goal of data analysis is the discovery of a model which fits the data. Statistical tools to accomplish this goal can differ in two ways: First, analysis tools differ in the kind of model which they fit to the data. For example, regression attempts to fit a linear subspace to the data points. Ordination attempts to fit a linear order to the data points. Clustering attempts to fit the data with a finite number of *clusters,* or subpopulations, each with distinct properties. We call this choice of model for an analytic tool its *model bias.* Second, analysis tools differ in the criteria used for goodness of fit. Regression typically seeks to minimize the sum of the squared distances of the data points from the regression subspace, but other measures, such as absolute value or a weighted sum, can be used. In clustering, the fitness criterion is usually the minimization of intracluster distance and simultaneous maximization of inter-cluster distance. The bias of the clustering procedure is then dependent on the distance function or metric used. We call this feature of an analysis tool its *fitness bias.*

We propose here a clustering methodology with a novel fitness bias. Our approach makes the clustering procedure easier to interpret and also leads to improved performance in some domains. Our rationale for the fitness bias is our concern for the uses of exploratory data analysis, and not an *a priori* judgement about similarity measures for data points. We assume that scientific data analysis is concerned with the patterns of cause and effect implicit in the data, and an appropriate analysis tool ought to be biased towards this in its model. In particular, a clustering methodology will attempt to find subpopulations of the data such that the observed data are highly contingent on the subpopulations. An optimal model of the data will be one which maximizes the *predictability* of data values, conditioned on the subpopulations. Our methodology thus maximizes the *utility* of the clustering, i.e., it attempts to minimize errors in predictions about samples from the data set. Further, we believe that it is particularly important in exploratory data analysis situations to *fit* the data without *distorting* the data, and our methodology therefore eschews all preprocessing of the data by, for example, normalization, substitutions for missing point-values, or elimination of outliers.

We take the distance metrics and functions, used in traditional clustering, to be ad hoc solutions to the problem of fitness bias, and inappropriate to most real world data analysis situations. Because we do not use distance functions, many of the problems of metric-based clustering do not arise. For instance, real scientific data sets are often heterogenous, or mixed, in their types. Some features of a data point may be categorical, others binary, and others real valued. To create a distance metric for such feature spaces introduces more ad hoc assumptions, or, worse, transforms the data to fit the analysis procedure. Secondly, incomplete data, i.e., data in which some or all points have *missing* values for some features, is common in real data sets. To use a distance metric on incomplete data requires some assumptions about the missing values, such as substitution of the mean, which again is a gross distortion of the original data. Thirdly, metric-based clustering cannot distinguish between important features, and those features in the data set which are noisy but which have no connection with the underlying cause and effect that determines the bulk of the other feature's values. Such "nuisance" features typically have to be filtered from the data set in advance of the clustering process. Finally, a measure of clustering *quality* is often not used, or is used separately from the clustering procedure itself. A clustering quality measure indicates not just which model fits "best," but provides some guidance on "how good" the fit actually is. This is critical, for instance, in deciding whether the data is better fit by two clusters, or by three clusters. Our methodology, however, incorporates a single measure of the *utility* of a clustering which 1) is meaningfully definable for continuous and discrete, ordered and unordered, feature types, 2) automatically ignores missing values in the data set, 3) automatically filters nuisance variables out of the eventual clustering, and 4) provides an integral measure of the quality of the clustering. Further, our approach is nonmetric (or, in statistical terms, nonparametric) in that our measures rely on the ordering of numeric data, but not the numeric distances.

We use our measure of utility, called *nonmetric fitness* (NMF) and described below in Section II-C, to guide a heuristic search over partitions of the data, seeking a global maximum. This approach is similar to conceptual clustering approaches to pattern recognition [1], in that the proposed usefulness of the clustering is an important factor in its fitness. Our approach is also similar Bayesian clustering [2], because we try to maximize the predictability of the actual data values, given the model. However, the system in [2] makes metric assumptions, that we do not, in assuming that the underlying distributions are multivariate Gaussian. Many of the tree-classifier systems [3]–[5] use fitness measures similar to NMF, but in classifier systems (using super-

Manuscript received August 15, 1989; revised August 23, 1990. This work was supported in part by the City of Bellingham, WA, and the Institute for Watershed Studies, Western Washington University.

The authors are with the Department of Computer Science, Western Washington University, Bellingham, WA 98225.

IEEE Log Number 9040947.

0162-8828/91/0200-0175$01.00 © 1991 IEEE

vised learning) instead of a clustering system (using unsupervised learning) such as ours.

Formal background for the approach is described below in Section II. Details of an implementation in the program, RIFFLE, are described in Section III. A comparison of its performance to that of the k-means clustering algorithm is described in Section IV-A, and some of the results of using the program on real world data are summarized in Section IV-B.

II. Formal Treatment

A. Clusterings

We assume the data constitute a set of I *points*, $D = \{x_i : i = 1, \cdots, I\}$ each of which is an ordered K-tuple, where K is the number of *features*, $x_i = \langle x_{i1}, \cdots, x_{iK} \rangle$. The features themselves will be named $P^1, \cdots, P^K$. The data can thus be viewed as a collection of I points in a K-dimensional space, the *feature space*. Each of the K features can be continuous or discrete, and may or may not have further structure, such as a natural zero or a natural unit (as in count data). Further, for each feature, "missing" or "unknown" can always be the value of a point.

A *clustering* of a data set D is a partition C of D into some number J of subsets, the *clusters* $C_1, \cdots, C_J$. The C_j are mutually exclusive and jointly exhaustive of D. Each data point $x_i \in D$ is given a number $j \in \{1, \cdots, J\}$ which is the number of the cluster to which x_i is assigned, and has no ordinal significance. We arbitrarily designate this feature, *cluster-number*, as the zeroth feature, so that the cluster-number for a data point x_i will be written x_{i0} and $x_{i0} = j$ iff $x_i \in C_j$. P^0 will then be another name for cluster-number.

B. Proportional Reduction in Error

We take the goal of a clustering to be accurate prediction of feature values for data points. We view cluster-number as simply another feature, and so we seek a quantitative measure of how well one or more features aid in the prediction of another. This is given by an estimate of the *reduction in error* achieved when using knowledge of the features, as opposed to prediction in ignorance. The measure we use is a generalization to an arbitrary number of features of Guttman's λ for two-dimensional cross-classification tables, which is extensively discussed in the literature [6]–[10]. The measure itself is only applicable to discrete features; our extension of it to clustering continuous features will be described in Section II-B-2.

1) Discrete Features: Consider the case where we are attempting to predict the value of one feature, P^3, on the basis of knowledge of two others, P^1 and P^2, and suppose that each of these features has three possible values, P^1 takes on values P_1^1, P_2^1, and P_3^1, and similarly for P^2 and P^3. With an adequate data set, we can obtain accurate frequency counts $f_{P_i^1 \wedge P_j^2 \wedge P_k^3}$ of the number of times a sample obtains values P_i^1, P_j^2, and P_k^3, for each i, j, and k. In other words,

$$f_{P_i^1 \wedge P_j^2 \wedge P_k^3} = |\{x : x_1 = P_i^1,\ x_2 = P_j^2,\ x_3 = P_k^3\}|.$$

See Fig. 1 where, for example, $f_{P_3^1 \wedge P_1^2 \wedge P_2^3} = 2$.

Now suppose for a particular data point x, we know $x_1 = P_3^1$ and $x_2 = P_1^2$, and wish to predict x_3. Clearly, we can do no better than look at all the frequency counts, for samples with the same values on P^1 and P^2, and choose the value of P^3 with the highest frequency. In other words, choose k such that $f_{P_3^1 \wedge P_1^2 \wedge P_k^3}$ is a maximum, which we denote: $\max_k \left(f_{P_3^1 \wedge P_1^2 \wedge P_k^3} \right)$. In Fig. 1, we have $f_{P_3^1 \wedge P_1^2 \wedge P_1^3} = 0$, $f_{P_3^1 \wedge P_1^2 \wedge P_2^3} = 2$, and $f_{P_3^1 \wedge P_1^2 \wedge P_3^3} = 5$, and so we should predict $x_3 = P_3^3$, and expect to be right about 5 out of 7 times.

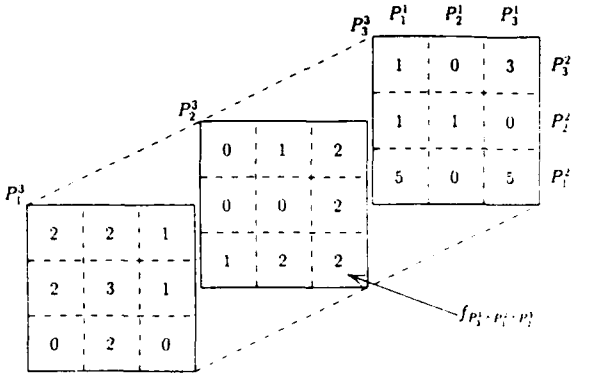


Fig. 1. A hypothetical frequency matrix for three features, P^1, P^2, and P^3, each with three possible discrete values. The frequency counts are entered in each cell, and the label for a typical cell illustrated.

If we make predictions for an entire collection of points, then our expected total correct percentage, in predicting P^3 on the basis of P^1 and P^2, would be

$$\text{Correct}(P^3 | \{P^1, P^2\}) = \frac{\Sigma_{i,j} \max_k \left(f_{P_i^1 \wedge P_j^2 \wedge P_k^3} \right)}{N}$$

where N is the total number of samples.

Generalizing to an arbitrary number of dimensions, an attempt to predict P^k, with values $P_{k'}^k$, conditioned on knowledge of a set of other features $\{P^i : i \in S\}$, $S \subseteq \{1 \cdots k-1, k+1, \cdots, K\}$, with values $P_{i'}^i$ is estimated to be correct with probability

$$\text{Correct}(P^k | \{P^i : i \in S\}) = \frac{\Sigma_{i'} \max_{k'} \left(f_{\left(\bigwedge_{i \in S} P_{i'}^i\right) \wedge P_{k'}^k} \right)}{N}.$$

If, on the other hand, we attempt to predict the value of a sample on a feature P^k using no information at all about the values of other features, then we can do no better than use the most common value of P^k, i.e., $P_{k'}^k$ where $\Sigma_{i'} f_{\left(\bigwedge_{i \in S} P_{i'}^i\right) \wedge P_{k'}^k}$ is a maximum, which we denote: $\max_{k'} \left(\Sigma_{i'} f_{\left(\bigwedge_{i \in S} P_{i'}^i\right) \wedge P_{k'}^k} \right)$. (The sums involved here are just the marginal totals of the frequency matrix.) If we use this for a guess, then our estimated probability of a correct prediction will be

$$\text{Correct}(P^k) = \frac{\max_{k'} \left(\Sigma_{i'} f_{\left(\bigwedge_{i \in S} P_{i'}^i\right) \wedge P_{k'}^k} \right)}{N}.$$

To obtain a measure of improvement based on these estimated probabilities, we can use the extent to which conditioning our predictions reduces error. The expected error rate in an unconditioned prediction is

$$\text{Error}(P^k) = 1 - \text{Correct}(P^k)$$

and the expected error rate in a conditional prediction is

$$\text{Error}(P^k | \{P^i : i \in S\}) = 1 - \text{Correct}(P^k | \{P^i : i \in S\}).$$

and the *proportional reduction in error* (PRE) is

$$\text{PRE}(P^k|\{P^i : i \in S\}) = \frac{\text{Error}(P^k) - \text{Error}(P^k|\{P^i : i \in S\})}{\text{Error}(P^k)}$$

$$= \frac{\Sigma_{i'}\left(\max_{k'}(f_{(\bigwedge_{i\in S} P^i_{i'})\wedge P^k_{k'}})\right) - \max_{k'}\left(\Sigma_{i'} f_{(\bigwedge_{i\in S} P^i_{i'})\wedge P^k_{k'}}\right)}{N - \max_{k'}\left(\Sigma_{i'} f_{(\bigwedge_{i\in S} P^i_{i'})\wedge P^k_{k'}}\right)}.$$

As a concrete example, we can calculate this quantity for the frequency matrix of Fig. 1, with the predicted feature, $k = 3$, and the known features $S = \{1, 2\}$ as follows:

$$N = 39$$
$$\sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_1} = 13$$
$$\sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_2} = 10$$
$$\sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_3} = 16$$
$$\max_k \sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_k} = 16$$
$$\text{Correct}(P^3) = 16/39 \approx 41\%$$
$$\max_k f_{P^1_1 \wedge P^2_1 \wedge P^3_k} = 5$$
$$\max_k f_{P^1_1 \wedge P^2_2 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_1 \wedge P^2_3 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_2 \wedge P^2_1 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_2 \wedge P^2_2 \wedge P^3_k} = 3$$
$$\max_k f_{P^1_2 \wedge P^2_3 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_3 \wedge P^2_1 \wedge P^3_k} = 5$$
$$\max_k f_{P^1_3 \wedge P^2_2 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_3 \wedge P^2_3 \wedge P^3_k} = 3$$
$$\sum_{i,j} \max_k f_{P^1_i \wedge P^2_j \wedge P^3_k} = 26$$
$$\text{Correct}(P^3|\{P^1, P^2\}) = 26/39 \approx 67\%$$
$$\text{PRE}(P^3|\{P^1, P^2\}) = \frac{26 - 16}{39 - 16} \approx 43\%.$$

In other words, the prediction of P^3 in ignorance will be correct 41% of the time, the prediction of P^3 using P^1 and P^2 will be correct 67% of the time, and so we can expect to be wrong 43% less often when we use P^1 and P^2 in the prediction of P^3 (assuming our sample is representative of the population).

If the set $\{P^i : i \in S\}$ contains only a single feature, then we write $\text{PRE}(P^k|P^i)$ for $\text{PRE}(P^k|\{P^i : i \in S\})$, and $\text{PRE}(P^k|P^i) = \lambda_{P^i P^k}$, Guttman's λ. Some properties of λ [6] are

1) λ lies between 0 and 1, inclusive, except when the entire population lies in a single cell of the table, in which case it is indeterminate.
2) λ is 1 if and only if all the population is in cells no two of which are in the same row or column.
3) If k and k' are independent, then λ is 0, but not necessarily vice versa.
4) λ is unchanged by permutations of rows or columns.

2) Continuous Features: To measure PRE on continuous features, discrete values are calculated from the continuous ones. The range of the feature is subdivided into J connected regions, and the discrete value of the continuous feature is the number of the region it falls into. This is justified by the observation that a clustering procedure, as a consequence of its model bias, produces only a finite number J of clusters. Even with a perfect clustering, feature predictions will be coarse, limited, for each feature, to one predicted value for each of the J clusters. On our model, we assume that each cluster will, accordingly, be associated with a single, connected subrange of each feature. For each J-clustering and for each continuous feature k, we choose $J - 1$ split-values, $s_{k_1} < \cdots < s_{k_{J-1}}$, and then define the discrete value for each sample x_i as

$$\text{discrete}_k(x_i) = j \text{ iff } s_{k_{j-1}} \le x_{ik} < s_{k_j}$$

where, for completeness, we can take $s_{k_0} = \min_i(x_{ik})$ and $s_{k_J} = \max_i(x_{ik}) + 1$. For example, with two clusters there will be a single split value and each data point will have either a "high" or a "low" value for each continuous feature. With three clusters, there would be "high," "medium," and "low" values. (More complex subsets could be imagined, but would greatly increase the complexity of the algorithm and, we believe, would find little use in practice.)

In any computation of PRE involving a continuous feature, it is understood that PRE is the maximum, over all such sets of split values, of the proportional reduction in error calculated in the usual way. Calculating such a maximum may involve a search over all candidate split values, or split values can be selected heuristically (as in our implementation, Section III), and these used as an approximation to the optimal split values.

C. A Nonmetric Measure of Clustering Fitness

The measure of error reduction PRE, defined above, can be used to define a measure of clustering fitness. The goodness of fit of a clustering is determined by how well feature values can be predicted using the clustering. Suppose, for example, we have a sample x, not part of the original data set, and we know only its feature values on the features in a given set, $\{P^i : i \in S\}$, and we want to predict a feature value x_j, with $j \notin S$. Using a given clustering of the data in this prediction is a two phase process. First, the cluster-number for x, i.e., x_0, is guessed, using the known feature values, and then the value of x_j is guessed, using x_0. The fitness of a clustering, therefore, can be measured by calculating $\text{PRE}(P^0|\{P^i : i \in S\})$ and $\text{PRE}(P^j|P^0)$. (We can, of

course, calculate $\text{PRE}(P^j|\{P^i : i \in S\})$ directly, but that answers a different question, regarding the intercorrelations of the features with each other. Here we seek an evaluation of the fitness of a clustering.) In a given clustering problem, we do not generally know j and S in advance, i.e., we do not know which features will be used in the prediction task. In fact, we take it to be part of the clustering task to determine which features *can* be used successfully in prediction. A data set, in other words, may be well clustered in some features, but also contain spurious or noisy features which have little relation to the clusters, and which could never be predicted accurately.

This leads to the following definitions. The *nonmetric fitness* (NMF^S) of a clustering C *in relation to* a feature set $\{P^i : i \in S\}$, is the average value of all terms of the form $\text{PRE}(P^0|\{P^i : i \in S'\})$ and of the form $\text{PRE}(P^j|P^0)$, where $S' \subseteq S$ and $j \in S$. A particular feature set (it need not be unique), for which NMF^S is a maximum is called an *optimal feature set* for C, and its nonmetric fitness is denoted simply by NMF. The fitness bias of our clustering methodology is toward clusterings with maximum NMF.

The introduction of the set S into our definition of clustering fitness, and the sets $S' \subseteq S$, permits further refinements in the notion of clustering fitness. Let the cardinality of S be $|S|$. If we restrict $|S|$ to be strictly less than the total number of features, $|S| < K$, we will obtain a clustering evaluated on a subset of features. Our fitness bias will then not only seek fit clusters, but will seek the best features for those clusters, resulting in "data reduction" on both the points (by grouping them into clusters) and on the features, but filtering out all but $|S|$ of them. On the other hand, if we restrict the size of S' in the definition of NMF we can control the amount of interdependence between features used to define the clustering. Setting $|S'| = 1$, for instance, requires the clustering to fit each feature in S independently of the others. Setting $|S'| = 2$ allows two-feature interactions, but excludes possible higher-order dependencies among features from consideration. (Both of these restrictions are provided as user options in our implementation, Section III.) The size of the optimal feature set $|S|$ is called the *number of significant features*, and the size of interactions allowed, $|S'|$, is called the *interaction-level*.

To illustrate the measure of clustering fitness, and as well the concomitant selection of split values to maximize NMF, consider the two-dimensional data of Fig. 2(a). We seek an optimal clustering into two clusters, with $|S| = 2$ (all features are significant), and $|S'| = 1$ (the interaction-level is one and we attempt to cluster on features independently). In Fig. 2(b) an optimal clustering and the two split values (dashed lines) are shown. The split values allow us to view the continuous features as discrete; each point will have either a "high" or "low" value on each feature, and consequently belong uniquely to one of the four cells of the frequency matrix. Points labeled 1 and 2 are clustered perfectly, because their value on any one feature P^0 (cluster-number), P^1 or P^2, determines the values on the other two. The point labeled "X", however, is more difficult. If it is assigned to cluster 1, then

$$\text{PRE}(P^0|P^2) = \text{PRE}(P^2|P^0) = 1.0$$

but

$$\text{PRE}(P^0|P^1) = \frac{(10+6)-11}{17-11} = 5/6$$

$$\text{PRE}(P^1|P^0) = \frac{(10+6)-10}{17-10} = 6/7.$$

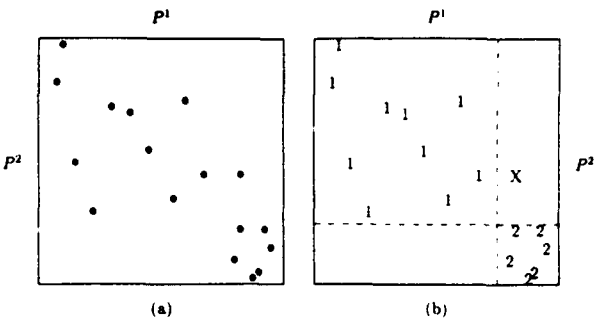


Fig. 2. An example data set (a) to be clustered using nonmetric fitness. Optimal clustering and split values are shown in (b). The point labeled "X" cannot be successfully clustered and will be assigned arbitrarily to cluster 1 or cluster 2.

On the other hand, if the point labeled "X" is assigned to cluster 2, then

$$\text{PRE}(P^0|P^1) = \text{PRE}(P^1|P^0) = 1.0$$

but

$$\text{PRE}(P^0|P^2) = \frac{(10+6)-10}{17-10} = 6/7$$

$$\text{PRE}(P^2|P^0) = \frac{(10+6)-11}{17-11} = 5/6.$$

In both cases, then, the NMF value will be $(1 + 1 + 6/7 + (5/6))/4 \approx 0.89$. The point labeled "X", therefore, can be assigned arbitrarily to either cluster, and both of the resulting clusterings are optimal. Any attempt to overcome this problem with the "X" point by adjusting the split values will create more problems than it solves, because more of the other points will then fall into one of the "troublesome" quadrants. This example also illustrates how maximization of PRE simultaneously on several different features, by adjusting their split values as well as the clustering, is necessary to achieve good fitness. Clustering one-dimensional, continuous-valued data on our criterion is a degenerate case, as any split value at all will give an NMF of 1.0 when cluster-numbers are selected to match discrete feature-values, and so we require $|S| \geq 2$.

III. Implementation

We have implemented our methodology in a computer program called RIFFLE, which is best described as a series of nested searches. The outermost loop searches for the best number of clusters ($J \geq 2$) simply by finding the best clustering for each number (in a user-specified range), and comparing the NMF values for each. One of the advantages of using NMF evaluations of clusterings is that fitness measures for clusterings with different J can be meaningfully compared. NMF is a measure of prediction accuracy, and whether one is predicting two values (high versus low) or three values (high, medium, or low), counts of correct and incorrect guesses can be compared (see Section IV-A-6, below).

The next level of search, given a fixed number of clusters J, is for the best cluster-numbering, i.e., assignment of points to clusters. Since a clustering is a partition of the points, the number of possible clusterings is $S(I, J)$ (Stirling numbers of the second kind, [11, pp. 90–91]), which prohibits exhaustive search. Instead, we begin with a random assignment of each

point to one of the J clusters, and then execute a hill-climbing search for improvements.

Currently this is done by reassigning a single point to a different cluster, recalculating NMF, and comparing the new fitness with the old. If any improvement is found, the point is left with its new cluster-number, otherwise the point is given its old cluster-number. In either case, other points are then examined to look for further improvement. Any time a point is successfully reassigned, *all* other points are then reexamined for possible further reassignment. This process continues until no improvements can be found by single-point reassignments, indicating we have reached a local maximum in NMF values. To avoid local maxima, the search may be repeated a number of times starting from a different initial random clustering. The number of repetitions necessary is, of course, domain dependent, but in practice we have never found more than about 50 to be necessary.

Nested within the search for optimal cluster-numbers is the evaluation of NMF, which involves a search for the optimal feature set and optimal split values for any continuous features in that set. User input relevant to this is the optimal feature set size $|S| = K^O$ and the interaction level $|S'| = K'$. If the interaction level is one, then for each feature P^k, we evaluate all terms of the form $\text{PRE}(P^k|P^0)$ and $\text{PRE}(P^0|P^k)$ and average these to give a "score" for P^k. If the interaction level is greater than one, all terms of the form $\text{PRE}(P^k|P^0)$ and $\text{PRE}(P^0|S')$ are computed, for all sets S' with $|S'| = K'$. Each feature P^k is then given a score by averaging all terms in which it appears, either as the predicted feature or as a member of the set S'. In either case, those K^O features with the highest scores are selected to form the optimal feature set, which in turn is used to compute NMF. (For $K' > 1$ this procedure is heuristic, and optimality is not guaranteed.)

Finally, nested within the search for optimal features and calculation of NMF, is the search for optimal split values for the continuous features. Although there are infinitely many sets of split values, there are only finitely many that make a difference to a given data set. If $J-1$ split values are sought for a total of I points, $J-1$ distinct points can be selected and their feature values used as the split values. An exhaustive search would therefore require examining $B(I, J-1)$ (binomial coefficient, I objects taken $J-1$ at a time) choices of points for split values. Currently, our implementation avoids this search by using another hill-climbing search. The data is sorted, in each feature, before the main loop of the procedure begins, so that initial split values can be selected at the quantiles of the data (medians for two clusters, quartiles for four clusters, etc). At each iteration, these values are adjusted up or down by one data point (in sorted order) and the NMF recalculated. If improvements are found, the new split values are retained, otherwise not.

The time complexity of our implementation, for a fixed number of clusters, can be computed as follows. Let

I = Number of points.
J = Number of clusters.
K = Number of features.
K' = Interaction level.
R = Number of repeated searches called for ≤ 50.
H = Average length of the hill-climbing search.
P = Number of PRE values to compute per NMF evaluation = $B(K, K')$.
Q = Time to compute each PRE value = $I \cdot J^{K'}$.
S = Time to sort feature scores = $K \log K$

Then the time complexity of our algorithm is on the order of $R \cdot H \cdot P \cdot Q + S$. For the most common case, interaction level $K' = 1$, and with $K \leq I$, this reduces to $O(H \cdot I \cdot J \cdot K)$. The size of H is difficult to predict, and in the worst case will be exponential (J^I), but in practice we have found the search to converge quickly to a local maximum. Letting the interaction level K' increase greatly increases the complexity, because of the large number of possible interactions among features, but we have found in practice that an interaction level of one works well even with dependent features (see Section IV).

The user input to the program consists of:

- The data.
- The number of features K.
- The type, continuous or discrete, of each feature.
- The minimum and maximum number of clusters to be examined.
- Optionally, the size of the optimal feature set. Default: the total number of features.
- Optionally, the interaction level. Default: one.
- Optionally, the number of times to repeat the search. Default: no repeats.

The user can request some or all of the following output, for each number of clusters between the input minimum and maximum, or for only the number of clusters with the best NMF:

- The cluster numbers for each point.
- The NMF value for the clustering.
- The features in the optimal feature set.
- The split values for each numeric feature in the optimal feature set.
- The PRE values for each feature individually with respect to the clustering.
- Means and variances for each cluster, for numeric features.

IV. Evaluation of RIFFLE's Performance

A. Monte Carlo Studies

In this section we compare the performance of RIFFLE to k-means clustering, a standard clustering procedure with good performance on Gaussian data. We compare their ability to recover clusters in data generated by Monte Carlo methods from two or more distinct distributions. We count the number of "correct" and "incorrect" classifications by the algorithms on the basis of the distributions that actually generated the points. Since the distributions have some degree of overlap, no procedure based solely on the data could correctly determine the originating distribution for every point, and so an "optimal algorithm" was used to obtain a lower bound on accuracy. Optimal clustering was done by assigning each data point to its *most likely* originating cluster, using the known distributions that generated the data points. The optimal algorithm therefore is not a clustering algorithm but serves only to obtain a lower bound on the misclassification rate. For the RIFFLE algorithm, the interaction-level was set to one, so that features were treated as independent. For the k-means algorithm, squared Euclidean distance was used.

1) Two-Dimensional Gaussian Clusters: In our first test, we generated two-dimensional Gaussian data in two subpopulations (similar to the data used in [12]), with subpopulation means along the $x = y$ diagonal, and at several different separations in means between the two subpopulations. The separation in means ranged from one to five times the standard deviation of each subpopulation about its own mean σ. (The two subpopu-

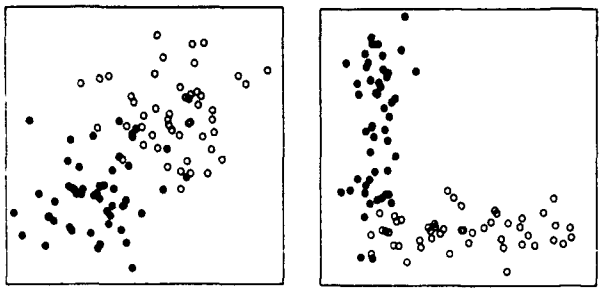


Fig. 3. Examples of synthetic two-dimensional data sets used in Monte Carlo tests. (a) Gaussian data and (b) boomerang data.

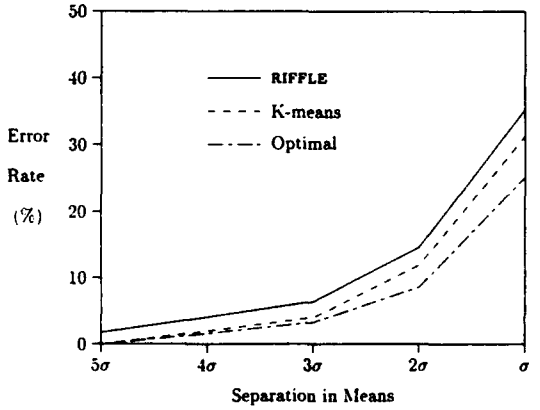


Fig. 4. Relative degradation of performance of RIFFLE, *k*-means, and optimal algorithms on two-dimensional Gaussian data. Errors increase for all three as the means of the two subpopulations are brought closer together. The separation in means is measured in terms of the standard deviation of each subpopulation about its mean.

lations each had the same variance.) One hundred points were generated in each experiment, with fifty in each cluster. For each parameterization the experiment was duplicated ten times, with different random number seeds, to obtain reasonable standard errors for the misclassification rates. A typical data set for this experiment is plotted in Fig. 3(a). Results for the RIFFLE and *k*-means algorithms, and the optimal reclassification scheme, are plotted in Fig. 4. In general, both algorithms performed well on Gaussian data.

2) Addition of Nuisance Features: A long-standing problem for many clustering algorithms [11, pp. 108–111] comes in the form of "cigar" shaped data, as illustrated in Fig. 5. Metric based clustering algorithms, which seek hyperellipsoidal clusters, typically break the cigars in half, as illustrated by the *k*-means clustering in Fig. 6. Clustering by RIFFLE, however, shown in Fig. 7, preserved the cigar shapes by placing more importance on the good fit of the clustering in two of the dimensions, and less importance on a poor fit in the third.

Data sets similar to the one in Fig. 5 were generated using the two-dimensional Gaussian data sets from the last section, with separation of means equal to 2σ. The cigar shape was created by introducing a third, "nuisance" feature, with values for the points randomly distributed over a range. The range of the nuisance feature varied from zero to four times the separation in means on the first two features. In Fig. 8 the performance of *k*-means is seen to degrade very severely as the range of nuisance noise increases. This is to be expected, since, as the nuisance

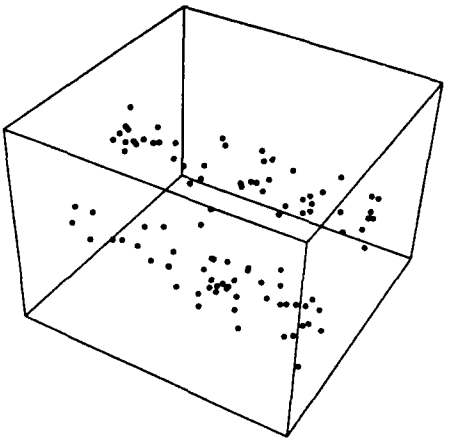


Fig. 5. Three-dimensional "cigar-shaped" data. In two dimensions the data are similar to Fig. 3(a). The points are randomly distributed in the third dimension.

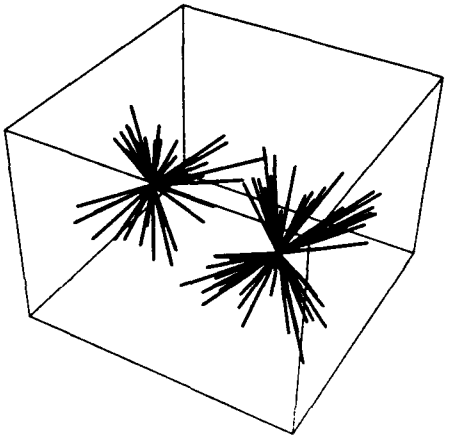


Fig. 6. Clustering of the cigar-shaped data from Fig. 5 by the *k*-means algorithm. Metric proximity dominates the clustering procedure, and the cigar-shaped structure is not recovered.

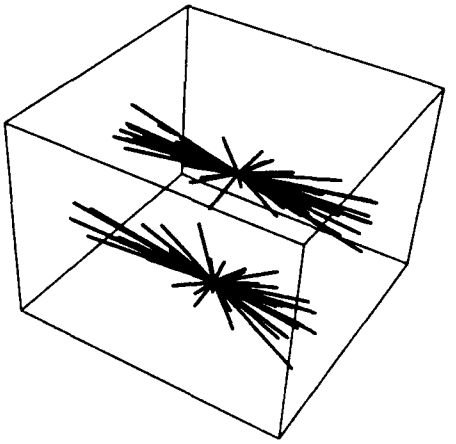


Fig. 7. Clustering of the cigar-shaped data from Fig. 5 by the RIFFLE algorithm. Because the indicated clustering fits well with two dimensions, the random third dimension is ignored.

feature increases in range, it dominates the other terms in the distance metric. The performance of RIFFLE, however, degrades

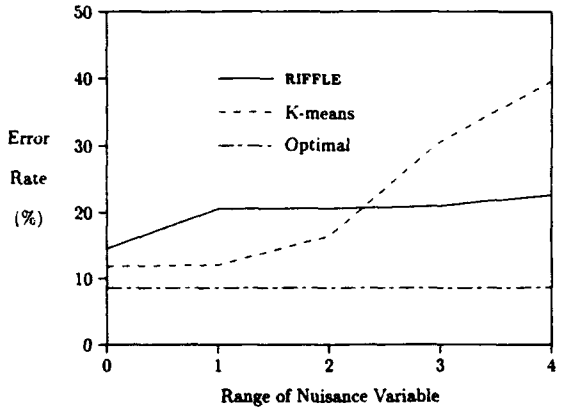


Fig. 8. Relative degradation of performance of RIFFLE, k-means, and optimal algorithms on three-dimensional, cigar-shaped data sets similar to that of Fig. 5. In these data sets the overlap in the first two dimensions was greater, and the range of the randomized third dimension was increased from zero to four times the separation in means in the first two dimensions.

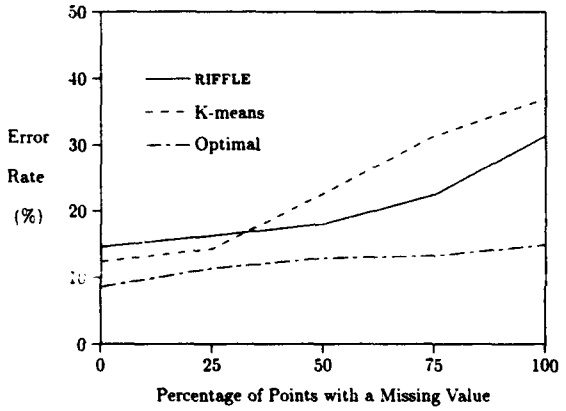


Fig. 9. Degradation of performance of RIFFLE, k-means, and optimal algorithms on two-dimensional Gaussian data similar to that in Fig. 3(a). The percentage of points which had one missing feature value was increased from zero to a hundred. The k-means algorithm required some distortion of the data in order to be usable: a substitution of the mean value for the missing values was used, and led to rapid degradation in performance. No preprocessing of the data was necessary for the RIFFLE procedure, and its degradation was less severe.

little because the fitness is nonmetric, and a good clustering in the first two features will dominate a clustering based primarily on the third feature, regardless of its range.

3) Incomplete Data: The same two-dimensional Gaussian data sets, with separation of means equal to 2σ, were also modified by taking a percentage of the points and marking one or the other of their two feature values as "missing." RIFFLE required no special treatment for missing values, since, with interaction-level one, each feature is examined independently of the others and a missing feature value is ignored in the calculation of PRE for that feature alone. However, since the standard k-means algorithm requires complete data sets, substitution of the mean value was used for missing values when running k-means. The performance of the two algorithms on this data is presented in Fig. 9. As the percentage of points with a missing value increases, the performance of k-means degrades more rapidly than RIFFLE.

4) Boomerang Data: Many data analysis situations involve clusters of points that are non-Gaussian. One common situation is when two populations represent different etiologies, but with

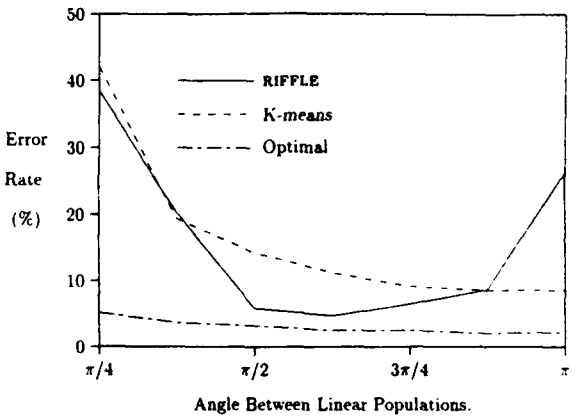


Fig. 10. Relative performance of RIFFLE, k-means, and optimal algorithms on two-dimensional boomerang data similar to that in Fig. 3(b). The angle between the two linear subpopulations varied from $\pi/4$ to π. RIFFLE's performance is equal or superior to that of k-means for most angles except the degenerate case of a straight line (π).

a common origin. They tend to cluster along two different linear subspaces of the feature space, resulting in "boomerang" shaped data, such as seen in Fig. 3(b). To simulate such data, two line segments were used. The "reference" line segment was selected parallel to the x-axis. Points from one cluster were scattered uniformly along this line, with added Gaussian noise in both the x and y dimensions. The second line segment was placed at several different angles to the first, from $\pi/4$ to π, and points from the second cluster were scattered uniformly along its length, with identical Gaussian noise in x and y. A typical data set for $\pi/2$ is shown in Fig. 3(b).

Error rates for k-means and RIFFLE on these data sets are plotted in Fig. 10. For angles close to $\pi/2$ RIFFLE outperformed k-means. The reason for this is that a distance metric clustering, forced to cluster into two groups, will usually lump most of the points at the "bend" of the boomerang into the same cluster. However, in a clustering by RIFFLE, split-values close to the bend are preferred because that gives each cluster a high PRE value on at least one feature, resulting in one "horizontal" cluster and one "vertical" cluster. In Fig. 11, clusterings by k-means (a) and RIFFLE (b) for a typical boomerang data set are shown. This figure may be compared to Fig. 3(b), where the "true" subpopulations for the points are given. If there is no marked difference between the linear trends of the clusters, however, as when the angle approaches zero or π, the performance of RIFFLE breaks down.

5) Categorical Data: Categorical data was simulated with various numbers of binary features. Two subpopulations were defined by randomly choosing a single, discrete probability value prob_k for each feature, giving the probability that a sample from subpopulation one would have a "0" value on that feature. The probability that a sample from subpopulation two would have a "0" was then set at $1 - \text{prob}_k$. The experiment was repeated for a number of features varying from 3 to 8. Results for both algorithms are in Fig. 12, where it can be seen that their performances are similar.

6) Recovering the Number of Clusters: While the "true" number of clusters in a data set is an ambiguous notion, we nevertheless attempted to assess RIFFLE's performance in this area with synthetic data sets similar to those in [11]. Three data sets were generated, one with strongly clustered points, one with weakly clustered points, and one with unclustered (randomly distributed) points. Points were scattered over the unit hypercube

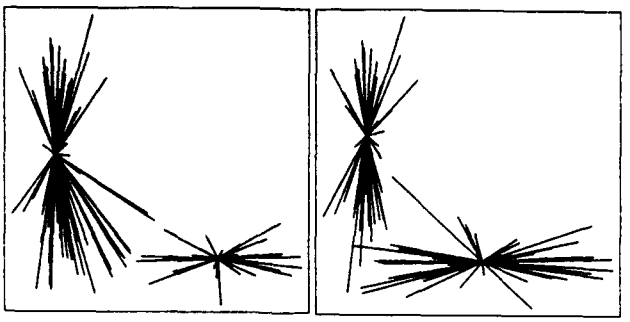


Fig. 11. Clusters generated by k-means (a) and RIFFLE (b) for boomerang data similar to that in Fig. 3(b). K-means clustering puts all the points at the "bend" in a single cluster because they are near each other in the metric. The RIFFLE clustering, however, separates the data into two subpopulations, each of which fits well with a particular dimension: one horizontal population and one vertical population.

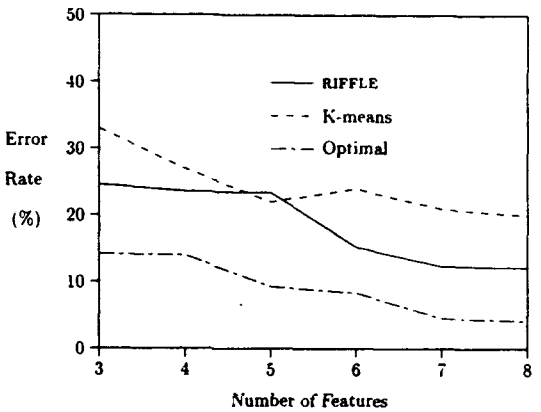


Fig. 12. Relative degradation of performance of RIFFLE, k-means, and optimal algorithms on binary categorical data. The number of binary features was varied from three to eight.

in five dimensions. For the clustered points, four subpopulation-centers were randomly selected. The strongly clustered points were normally scattered about these centers with a standard deviation of 0.01 in each dimension (no covariance); the weakly clustered points were scattered about the same centers with a standard deviation of 0.1 in each dimension. These three data sets were each clustered into two to twelve clusters by RIFFLE, and the resulting fitness values are plotted in Fig. 13. For the strongly clustered data, a clear peak is seen at the correct number, four, while for the weakly clustered data, a slight peak is still seen at the correct number. This compares well to the Davies–Bouldin index and the modified Hubert Γ index, which are plotted for similar data sets in [11, pp. 186–188]; both of these indices indicated four clusters in the strongly clustered data, but showed a slight preference for three clusters in the weakly clustered data.

In the plot for random data in Fig. 13, a tendency toward better fitness values for larger numbers of clusters can be observed. This is the well known problem of "over-fitting" a model, which plagues all data analysis situations. If necessary, a penalty for larger numbers of clusters could be introduced, perhaps along lines suggested in [13], but we have not found this necessary in practice.

B. Real World Data

1) Known Clusters: We presented two real world data sets with known properties to RIFFLE and to the k-means algorithm, to see if they could recognize the originating subpopulations. The first was Fishers's "iris" data [14], consisting of two sepal and two petal measurements from 150 irises, 50 from each of three species. We first attempted to recover the "true" number of clusters from the data. The NMF fitness values for each number of clusters from two to twelve are plotted in Fig. 14, which shows a clear peak at three. This compares favorably to the modified Hubert Γ index [15] and the fuzzy hypervolume and density indexes [16], which have been tested on the iris data, and which indicate three clusters. The Davies–Bouldin index, however, does not seem to indicate a preference for any number [15]. Using three as the correct number of clusters, we compared RIFFLE and k-means clustering. Each correctly reclassified 134 out of 150, or 88% of the irises.

The second real world data set was the "8OX" data set from [11], consisting of eight features extracted from 45 handwritten characters, 15 each of "8", "O", and "X". Fitness values for this data are also plotted in Fig. 14, but they do not reveal a clear peak. We believe this is due to the small number of data points in the 8OX data, and the fact that the clusters in the 8OX data are not well separated. Assuming, however, that each character represents a "true" cluster, we compared RIFFLE and k-means clustering on this data set. RIFFLE correctly reclassified 37 out of 45, or 82%, of the characters while k-means correctly reclassified only 30 out of 45, or 67%.

2) Unknown Clusters: In collaboration with colleagues, we are also applying RIFFLE to many ongoing problems in the analysis of real world data sets. In each case the program has

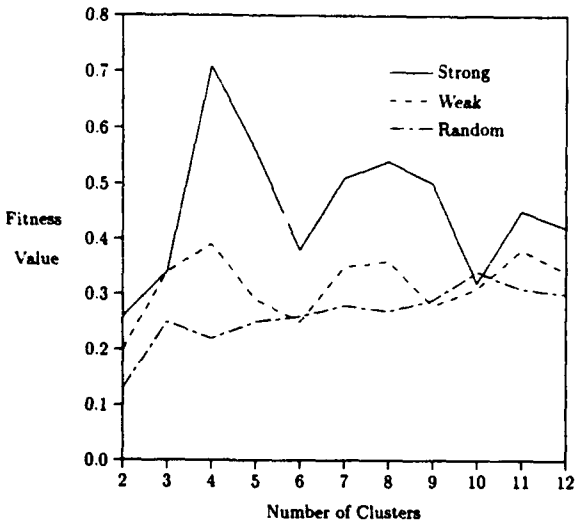


Fig. 13. Fitness values generated by RIFFLE for synthetic data with five features. Data which was strongly or weakly grouped into four subpopulations show a peak in fitness at four. Random data do not result in such a peak.

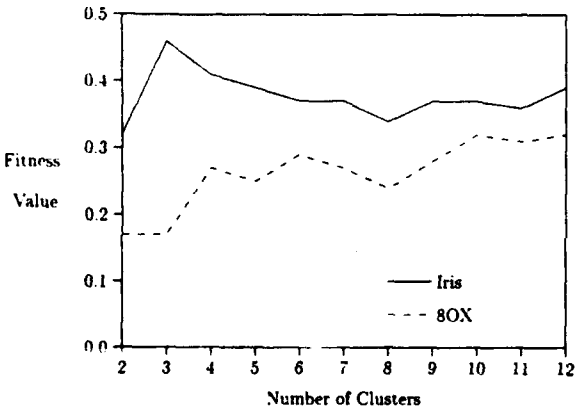


Fig. 14. Fitness values for the iris data set and the 8OX data set, for varying number of clusters, as determined by RIFFLE. For the iris data, a peak is seen at three. For the 8OX data set, no clear peak was identified, indicating that the data are not as well clustered.

created "meaningful" clusters, in some cases revealing previously unsuspected facets of the data to experts. In a year-round ecological study of a northwestern monomictic lake [17], [18], RIFFLE meaningfully clustered both the physical–chemical features and the phytoplankton species data. The physical–chemical data were separated into epilimnion, hypolimnion, and thermocline samples, even though data points were collected from three basins of the lake with quite dissimilar physical characteristics, and throughout the year. The phytoplankton samples were separated into summer versus winter samples, as these were the most dissimilar populations, with a clear break at fall turnover. Further, rare species, with low variance relative to the rest of the data set but with a high degree of association to the common algal blooms, were identified as optimal features. All other analysis tools used on the data failed to accomplish this. In another data set, gathered as part of the national acid rain survey [19] and involving hundreds of lakes, RIFFLE successfully partitioned lake samples into "impacted" and "not impacted" clusters. In a third data set, dealing with nonpoint-source pollution of an urban stream [20], RIFFLE was able to partition the samples into "polluted" and "unpolluted" clusters based solely on data involving counts of macroinvertebrates found at the sites, regardless of season. Again, other analysis tools failed to do this.

V. Conclusion

We have proposed an approach to clustering based on the principle that clusters should be selected to maximize their actual utility in predicting feature values, not ad hoc measures of similarity in feature space. We have defined a quantitative measure of this utility, called *nonmetric fitness*, which 1) is applicable to both discrete and continuous features, 2) can automatically ignore some or all noisy but irrelevant features, 3) can cluster incomplete data without assumptions about the missing values, and 4) provides some guidance in regard to the correct number of clusters. We have also implemented a clustering procedure, RIFFLE, using nonmetric fitness, and tested it on synthetic and real world data. We compared the performance of RIFFLE to k-means clustering and illustrated several cases where RIFFLE was superior. We are currently using RIFFLE, in collaboration with domain experts, in exploratory data analysis on real world problems, where it has proven a valuable adjunct to traditional statistical tools. We hope that our work will stimulate the creation of other clustering algorithms based on such fitness measures, as well as the use of these measures in other disciplines and data analysis tools.

References

[1] R. S. Michalski and R. E. Stepp, "Learning from observation: Conceptual clustering," in *Machine Learning, An Artificial Intelligence Approach*. Los Altos, CA: Morgan Kaufman, 1983, pp. 331–363.
[2] P. Cheeseman, J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman, "Autoclass: A Bayesian classification system," in *Proc. Fifth Int. Conf. Machine Learning*. Los Altos, CA: Morgan Kaufmann, 1988.
[3] L. Breiman, J. H. Freidman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
[4] E. G. Henrichon, Jr. and K. S. Fu, "A nonparametric partitioning procedure for pattern classification," *IEEE Trans. Comput.*, vol. C-18, pp. 614–624, 1969.
[5] J. A. Sonquist, E. L. Baker, and J. N. Morgan, *Searching for Structure*, Review ed. Ann Arbor, MI: Inst. Social Res., Univ. Michigan, 1973.

[6] L. A. Goodman and W. H. Kruskal, "Measures of association for cross classifications," *J. Amer. Statist. Assoc.*, vol. 49, pp. 732–764, 1954.
[7] ——, "Measures of association for cross classifications ii: Further discussion and references," *J. Amer. Statist. Assoc.*, vol. 54, pp. 123–163, 1959.
[8] ——, "Measures of association for cross classifications iii: Approximate sampling theory," *J. Amer. Statist. Assoc.*, vol. 58, pp. 310–364, 1963.
[9] ——, "Measures of association for cross classifications iv: Simplification of asymptotic variances," *J. Amer. Statist. Assoc.*, vol. 67, pp. 415–421, 1972.
[10] L. Guttman, *An Outline of the Statistical Theory of Prediction.* New York: S.S.R.C, 1941.
[11] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[12] C. K. Bayne, J. J. Beauchamp, C. L. Begovich, and V. E. Kane, "Monte Carlo comparisons of selected clustering procedures," *Pattern Recog.*, vol. 12, pp. 51–62, 1980.
[13] J. Rissanen, "Stochastic complexity and modeling," *Ann. Statist.*, vol. 14, no. 3, pp. 1080–1100, 1986.
[14] R. A. Fisher, "Multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.
[15] R. C. Dubes, "How many clusters are best?—An experiment," *Pattern Recog.*, vol. 20, no. 6, pp. 645–663, 1987.
[16] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 773–781, 1989.
[17] W. J. Ehinger, "Phytoplankton composition and temporal variation among the three basins of Lake Whatcom, Washington," Master's thesis, Western Washington Univ., Bellingham, WA, 1988.
[18] G. B. Matthews, "Clustering heterogeneous ecological data," in *Annu. Conf. Int. Soc. Ecological Modeling*, Davis, CA, 1988.
[19] R. A. Linthurst, D. H. Landers, J. M. Eilers, D. F. Brakke, W. S. Overton, E. P. Meier, and R. E. Crowe, *Characteristics of Lakes in the Eastern United State, Volume I. Population Descriptions and PhysicoChemical Relationships.* Washington, DC: U.S. Environmental Protection Agency, No. EPA/600/4-86/007a, 1986.
[20] R. A. Matthews, G. B. Matthews, and B. Hachmöller, "Ordination of benthic macroinvertebrates along a longitudinal stream gradient," in *Annu. Conf. North American Benthological Society*, Blacksburg, VA, 1990.

Geoffrey Matthews received the B.A. degree, cum laude, in philosophy from the University of California, Riverside, in 1973, and the Ph.D. degree, cum laude, in the history and philosophy of science from Indiana University in 1979.

His graduate and postdoctoral work concerned philosophical problems in the foundations of statistical physics, particularly the relations between entropy, information, causality, and the arrow of time. From 1984 to 1985 he worked at Hughes Aircraft Corporation, developing artificial intelligence applications in support of software engineering. Since 1985 he has been a Professor of Computer Science at Western Washington University, Bellingham. His research interests include pattern recognition, robust data analysis, natural language processing, and functional programming languages.

Dr. Matthews is a member of the Association for Computing Machinery and the IEEE Computer Society.

James Hearne received the B.A., M.A., and Ph.D. degrees in philosophy from the University of California, Riverside.

He has taught both Philosophy and Computer Science at several universities, including California State University at San Bernardino and Fullerton and the University of California at Riverside. He was also employed at Hughes Aircraft Company and Rockwell International where he worked on applications of artificial intelligence technology to software development environments and sonar signal processing. Since 1986 he has been an Assistant Professor of Computer Science at Western Washington University, Bellingham, where he has pursued research in exploratory data analysis and the possible worlds paradigm for distributed computing.

Reprinted from

Canadian Journal of Fisheries and Aquatic Sciences

Réimpression du

Journal canadien des sciences halieutiques et aquatiques

Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream

G. B. MATTHEWS, R. A. MATTHEWS, AND B. HACHMÖLLER

Volume 48 • Number 11 • 1991

Pages 2184–2190

Canada

Fisheries and Oceans Pêches et Océans

Printed in Canada by The Runge Press Limited
Imprimé au Canada par The Runge Press Limited

Mathematical Analysis of Temporal and Spatial Trends in the Benthic Macroinvertebrate Communities of a Small Stream

G. B. Matthews

Computer Science Department, Western Washington University, Bellingham, WA 98225, USA

R. A. Matthews

Huxley College of Environmental Studies, Western Washington University, Bellingham, WA 98225, USA

and B. Hachmöller

Barfußerstraße 42, 3550 Marburg, Germany

Matthews, G. B., R. A. Matthews, and B. Hachmöller. 1991. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. Can. J. Fish. Aquat. Sci. 48: 2184–2190.

Macroinvertebrates were collected at four sites in Padden Creek, a small second-order stream in Whatcom County, Washington, USA. Two upstream sites were characterized by high densities of sensitive taxa, predominantly mayflies, stoneflies, and caddisflies, and two downstream sites showed high densities of tolerant taxa, especially true flies, annelids, *Baetis* mayflies, and gastropods. Despite the small sample size, some statistical techniques proved useful. The first two components of correspondence analysis were used to confirm the existence of both seasonal and spatial trends in the benthic macroinvertebrate populations of the stream. Neither component alone, however, ordinated the samples with respect to these trends. Combinations of the first two components were required. A standard clustering technique, *k*-means clustering with squared Euclidean distance, further confirmed the seasonal trend. Nonmetric clustering, not widely used in the analysis of ecological data, was necessary to confirm the spatial trend. Nonmetric clustering was also able to identify a small number of "significant" taxa, i.e. taxa that reliably served as indicators of spatial position on the stream.

On a effectué un échantillonnage des macroinvertébrés à quatre sites du ruisseau Padden, un petit cours d'eau de second ordre situé dans le comté Whatcom de l'État de Washington (É-U). Des densités élevées de taxons sensibles étaient caractéristiques des deux sites d'amont, en particulier des éphémères, des perles et des phryganes, tandis que les deux sites d'aval abritaient des densités élevées de taxons tolérants, surtout des mouches, des annélides, des éphémères du genre *Baetis* et des gastéropodes. Malgré la faible taille des échantillons, certaines méthodes statistiques se sont révélées utiles. Ainsi, les deux premières composantes de l'analyse factorielle de correspondance ont permis de confirmer l'existence de tendances saisonnières et spatiales dans les populations de macroinvertébrés benthiques du cours d'eau. Toutefois, ni l'une ni l'autre de ces composantes n'a permis d'effectuer une ordination des échantillons en ce qui concerne ces tendances, ordination obtenue toutefois par la combinaison des deux premières composantes. L'agglomération de moyennes *k* couplée à la distance euclidienne au carré, une technique agglomérative normalisée, a permis d'étayer cette tendance saisonnière. L'agglomération non métrique, rarement utilisée dans l'analyse de données écologiques, a été nécessaire pour confirmer la tendance spatiale. Cette dernière analyse a aussi permis d'identifier un faible nombre de taxons "significatifs", c'est-à-dire des taxons qui ont servi d'indicateurs fiables de la position spatiale dans le cours d'eau.

Received September 9, 1990
Accepted May 2, 1991
(JA718)

Reçu le 9 septembre 1990
Accepté le 2 mai 1991

One of the fundamental principles of mathematical ecology is that changes in the statistical makeup of the biota are reflections of changes in the physical environment. The dominance of certain taxa at a particular site or between sites can serve as a quantifiable record of the strength and direction of environmental changes (Faith and Norris 1989). In the ecology of streams, there are often two dominant environmental changes, one associated with time and the other with location (Green 1974). The benthic community varies with the season, and also with its spatial position in the stream. Many benthic macroinvertebrates have habitat requirements that correspond to longitudinal gradients, upstream to downstream. For example, because they require highly oxygenated waters, many stoneflies are restricted to headwater streams, which are often less polluted and more turbulent than downstream reaches (Hynes 1970; McCafferty 1981). Many other stream characteristics can be viewed as changing along this longitudinal gradient, due to the unidirectional downstream flow. This view of streams as gradients has influenced many of the fundamental theories on how streams function, including organic matter processing, macroinvertebrate community trophic structure, in-stream primary productivity, and nutrient cycling (see Minshall 1988 and Fisher 1983 for general reviews). However, the complex distributions and patterns exhibited by macroinvertebrates

make statistical confirmation of such relationships difficult. The problem of identifying reliable taxonomic indicators of environmental changes is even more difficult.

In this paper we used ordination by correspondence analysis and clustering by two techniques, *k*-means clustering and nonmetric clustering, to obtain statistical confirmation of the benthic macroinvertebrate response to both the longitudinal and the seasonal trends. Correspondence analysis is well documented in the literature (e.g. Gauch et al. 1977; Kenkel and Orloci 1986; ter Braak 1986), but, while ordination has been used extensively for finding and confirming terrestrial vegetation gradients (e.g. Minchin 1987), it has been used much less frequently to examine gradients in stream data (e.g. Green 1974; Culp and Davies 1980; Sheldon and Haick 1981; Schaeffer and Perry 1986; Faith and Norris 1989). *K*-means clustering is also widely used in many fields (Jain and Dubes 1988). Nonmetric clustering, described in the Appendix, is a new technique and has not been widely applied to ecological data although we have found it useful in a variety of aplications (Matthews and Hearne 1991; Mathews et al. 1991). We found that the combination of these three analytical techniques provided an excellent approach to our data set. The spatial and temporal trends were both revealed by correspondence analysis. The temporal trend was confirmed by *k*-means clustering which successfully separated samples by date, and the spatial trend was similarly confirmed by the nonmetric clustering which successfully separated samples by site.

The data we used for our analyses were collected from Padden Creek, a small second-order stream located adjacent to the city of Bellingham in Whatcom County, Washington. Hachmöller (1989) and Hachmöller et al. (1990) found that the macroinvertebrate fauna in Padden Creek showed distinct upstream and downstream distribution patterns. These distribution patterns were thought to be related to differences in the riparian community, especially canopy cover, and the input of nonpoint-source runoff from residential and agricultural areas, which created a turbid, nutrient-enriched "lower reach" in the creek.

Methods

Macronvertebrate Sampling

Four sites were sampled in Padden Creek (Fig. 1). Site 1 was located approximately 1 km downstream from the Lake Padden outfall in a forested, relatively undisturbed area. Site 2 was located in a channelized reach that had a less diverse substrate than Site 1. Both Sites 1 and 2 were upstream from the confluence of Padden and Connelly Creeks. Connelly Creek is a nutrient-enriched tributary that drains agricultural and residential lands. Site 3 was located about 1.5 km downstream from Connelly Creek in a forested city park that was more disturbed than Site 1. Site 4 was located in a freshwater wetland close to the mouth of Padden Creek. Based on vegetation, water quality, and substrate sampling, Hachmöller et al. (1990) and Uhlig (1991) characterized the four sites as in Table 1.

The macroinvertebrate samples were collected monthly at each site from June through October 1988 using a Surber sampler (1-mm net mesh). Ten samples were collected at each site on each date. The invertebrates were keyed to the lowest practical taxon (genus in most cases) using the following references: Anderson (1976), Edmunds and Jensen (1976), Hatch (1953–65), Jewett (1959), Merritt and Cummins (1984), Pennak (1978), Ricker and Scudder (1975), Ross (1937), Stark and Gaufin (1976), and Stone et al. (1965). Macroinvertebrate densities for each taxon were calculated as the average number of individuals per square metre ($n = 10$ per site and date).

Statistical Tests

Throughout this section, a "sample" refers to the pooled macroinvertebrate densities at a unique site and date; there were 20 samples in this study (4 sites × 5 dates). Individual macroinvertebrate densities for each taxon are called "replicates." There were 10 replicates for each taxon (63 taxa) at each date and time (a maximum of 12 600 replicates, many of which had values of zero). Some statistical tests were performed on both the sample data averaged by replicate and the raw data, not averaged by replicate; however, only the results from the averaged sample tests are reported here. Generally, as might be expected, the raw data yielded similar results, but with larger variances.

We ordinated the samples using correspondence analysis. Correspondence analysis (also called reciprocal averaging) determines taxa scores and sample scores in an "uninformed" manner, i.e. without prior grouping of the samples. Thus, samples are ordinated independently of information regarding the actual site or date at which they were collected. For our purposes, it was important that the correspondence analysis procedure give several ordinations of the samples (first, second, third components, etc.), for we found that two components were necessary to reveal trends indicated by our subjective evaluations. The correspondence analysis procedure is similar to principal components and factor analysis, but has been shown to be superior to these methods in typical environmental data sets (Kenkel and Orloci 1986; Ludwig and Reynolds 1988).

The data were also clustered by the *k*-means algorithm using squared Euclidean distance, and nonmetric clustering. *K*-means clustering (Jain and Dubes 1988) views the samples as points in *n*-dimensional space, where *n* is the number of taxa. It seeks "clusters" of samples such that the distance between samples from the same cluster is generally less than the distance between samples from different clusters. The measure of distance between samples is called the metric. A clustering is optimal in the metric sense if it maximizes the difference between the average intracluster distance and the average intercluster distance. There are many measures of "distance" for samples, and the choice of a particular distance metric can have a radical effect on the resulting clusters. For our *k*-means clustering we used squared Euclidean distance. Nonmetric clustering, described in the Appendix, is a new procedure that does not use a distance metric to determine clusters (Matthews and Hearne 1991). Instead, a clustering is optimal in the nonmetric sense if it maximizes the association between clusters and a large number of taxa. Each taxon is also given a "score" by nonmetric clustering, which is a measure of how strongly that particular taxon is associated with the clustering. Both nonmetric and *k*-means clustering are uninformed procedures, like correspondence analysis, and do not require prior grouping of samples.

Correspondence analysis, metric clustering, and nonmetric clustering were also used in an effort to identify diagnostic taxa, i.e. a subset of the taxa that could be used as indicators of environmental conditions. Correspondence analysis not only ordinates the samples, but also ordinates the taxa, and thus "large" taxa scores might be taken to indicate taxa important

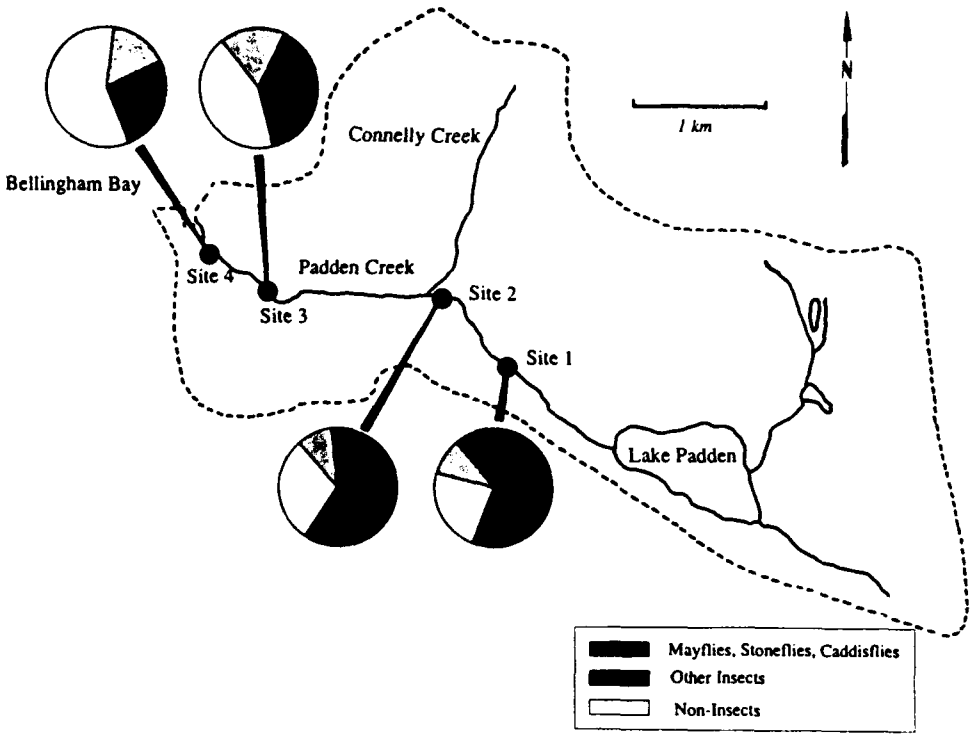


FIG. 1. Padden Creek sampling sites and relative proportions of major macroinvertebrate taxa. The macroinvertebrate proportions were calculated by averaging the macroinvertebrate densities (no./m^2) at each site for the entire study period.

TABLE 1. Characterization of the four sampling sites.

Factor	Site 1	Site 2	Site 3	Site 4
Nutrient concentration	Low–moderate	Low–moderate	Elevated	Elevated
Riparian vegetation	Second-growth coniferous forest	Alder; gaps in canopy	Second-growth coniferous forest	Freshwater wetland
Stream gradient	61 m/km	19 m/km	8 m/km	11 m/km
Substrate	Diverse cobble–pebble	Uniform cobble–pebble	Diverse pebble–sand	Diverse pebble–sand

to the correspondence analysis ordination. K-means clustering does not rank taxa in importance, and so was not used to identify diagnostic taxa. Nonmetric clustering, however, is designed to cluster data and simultaneously identify the taxa that are "important" with respect to these clusters (Matthews and Hearne 1991). In this regard it is similar to conceptual clustering techniques (Fisher and Langley 1986), which not only cluster the data, but attempt to show how those clusters can be characterized by a small subset of the data parameters. A nonmetric clustering which is meaningfully related to a spatial or longitudinal trend will also give a list of important taxa, which could be used as indicators of that trend.

Results

Hachmöller (1989) and Hachmöller et al. (1990) found that the most abrupt change in macroinvertebrate community structure occurred between Sites 2 and 3, which was attributed primarily to the influence of Connelly Creek. These changes can be seen in the pie charts summarizing the benthic community in Fig. 1. Mayflies, stoneflies, and caddisflies were collected in greater densities at the upstream sites (Sites 1 and 2); these three orders made up 62–67% of the macroinvertebrate densities at the upstream sites, but only 26–40% of the densities at the downstream sites (Sites 3 and 4). In addition, many of the uncommon taxa (less than 0.5% of the total density) were collected more frequently at the upstream sites, especially large, predatory stoneflies. This may be an artifact of the taxonomic technique because not all taxa were identified to the same level. In particular, Chironomidae and many of the noninsect taxa were identified only to family. This is a pervasive taxonomic dilemma, and its relevance to our statistical tests will be discussed below. In general, the macroinvertebrates collected at the downstream sites were mostly taxa having relatively cosmopolitan distributions such as *Baetis* and Chironomidae and included a large proportion of noninsect taxa such as oligochaetes, gastropods, etc.

Table 2 lists the average densities (number per square metre) for the most common taxa (greater than 0.5% of the total density) that were collected from Padden Creek from June through October 1988. A complete listing of the 63 Padden Creek taxa is given in Hachmöller (1989). It should be noted that the des-

TABLE 2. Macroinvertebrate densities and nonmetric clustering (NMC) scores for major taxa.

Padden Creek macroinvertebrate taxa	% total density	Average densities (no./m²) Site 1	Site 2	Site 3	Site 4	NMC score
Plecoptera						
Malenka spp.	5.0	78.57	80.94	3.44	6.45	
Skwala spp.	0.6	3.87	11.84	1.72	2.15	0.47
Suwallia/Triznaka/Sweltsa complex	2.9	75.77	19.59	0.21	1.50	0.89
Ephemeroptera						
Baetis spp.	10.3	48.43	130.88	60.06	108.93	0.12
Cinygmula spp.	2.4	46.71	32.29	1.93	1.29	0.68
Epeorus spp.	3.9	95.58	33.58	1.29	0.21	
Ironodes spp.	2.3	36.38	32.50	4.73	3.01	0.47
Paraleptophlebia spp.	3.7	13.56	41.11	40.04	31.00	
Serratella spp.	1.9	2.36	57.04	4.52	1.50	
Trichoptera						
Glossosoma spp.	5.4	31.43	119.04	25.61	7.10	
Hydropsyche spp.	4.5	14.52	29.06	8.82	1.07	0.33
Rhyacophila spp.	1.1	24.54	10.33	1.29	0.64	0.89
Parapsyche spp.	4.8	8.18	12.70	107.63	33.15	−0.47
Diptera						
Chironomidae	7.7	60.70	87.83	59.20	54.03	0.26
Simuliidae	4.0	17.00	0.21	53.17	64.79	−0.33
Amphipoda						
Gammarus lacustris	0.7	0.00	0.86	5.59	17.43	−0.80
Annelida						
Enchytraeidae	31.4	210.97	260.05	237.88	355.42	−0.26
Lumbriculidae	2.2	3.44	9.25	23.03	39.61	−0.68
Gastropoda						
Ferissia	0.5	0.00	3.87	10.97	2.79	−0.41
Gyraulus	1.4	0.00	13.99	9.68	24.54	−0.26

ignation of "common" is somewhat arbitrary because, again, not all taxa were identified to the same level.

Confirmation of the observed longitudinal and seasonal trends by correspondence analysis can be seen in Fig. 2, which plots all samples by the first two components of correspondence analysis. Neither trend, however, corresponds well with a single component of correspondence analysis. Instead, the seasonal differences tend to spread along a "northwest–southeast" line, and the longitudinal trends spread along an orthogonal, north-east – southwest" line. We believe that this observation is important, as the emphasis in much statistical ecology is on recognizing a single, dominant gradient in the population. This is the motivation behind "detrended" correspondence analysis, for example, which attempts to force a one-dimensional ordination for data sets. In our case, a two-dimensional ordination was essential.

The ordinations by correspondence analysis led to difficulties in the identification of indicator taxa. First, as seen in Fig. 2, neither of the first two sample score components, alone, corresponds with the trends of interest. Each is a combination of both trends. Accordingly, neither of the first two taxa scores could be used to determine indicator taxa for either trend. Second, although correspondence analysis taxa scores were partially associated with the trends (for example, positive taxa scores were generally assigned to "upstream" taxa and negative taxa scores to "downstream" taxa) the correspondence analysis scores were strongly influenced by rare taxa. Only three of the top 20 correspondence analysis taxa scores were from common taxa, these three being *Hydropsyche*, *Malenka*, and *Serratella*.

The seasonal trend was confirmed by *k*-means clustering, which separated the samples by date. The June and July samples were placed in one cluster and the August, September, and October samples in the other, except for one August sample which was placed in with the June and July samples (see Fig. 2a). On the other hand, nonmetric clustering confirmed the observed longitudinal trend, and clustered all upstream (Sites 1 and 2) samples into one cluster and all downstream (Sites 3 and 4) samples into the other cluster (see Fig. 2b).

The attempt to identify indicator taxa using nonmetric clustering was very successful. Unlike correspondence analysis, most of the top taxa scores produced by nonmetric clustering were from common taxa. These 15 out of the 20 top scores were common taxa, and are listed in Table 2. This was impressive considering there were only 20 common taxa and that nonmetric clustering is "naive" in that it did not use total macroinvertebrate density as a selection criterion. Further, we verified the robustness of this taxonomic subset using a "leave-one-out" strategy. The nonmetric clustering taxa scores were recalculated based on only 19 "training" samples, leaving one sample out, and then the group (upstream or downstream) for the omitted sample was predicted using taxa scores generated from the other 19 samples. This procedure was repeated with each sample being the one omitted, obtaining 20 tests; thus we obtained an estimate of the rate at which errors might occur in using these taxa scores to classify unknown samples, by simply counting the number of the "left-out" samples that were mis-

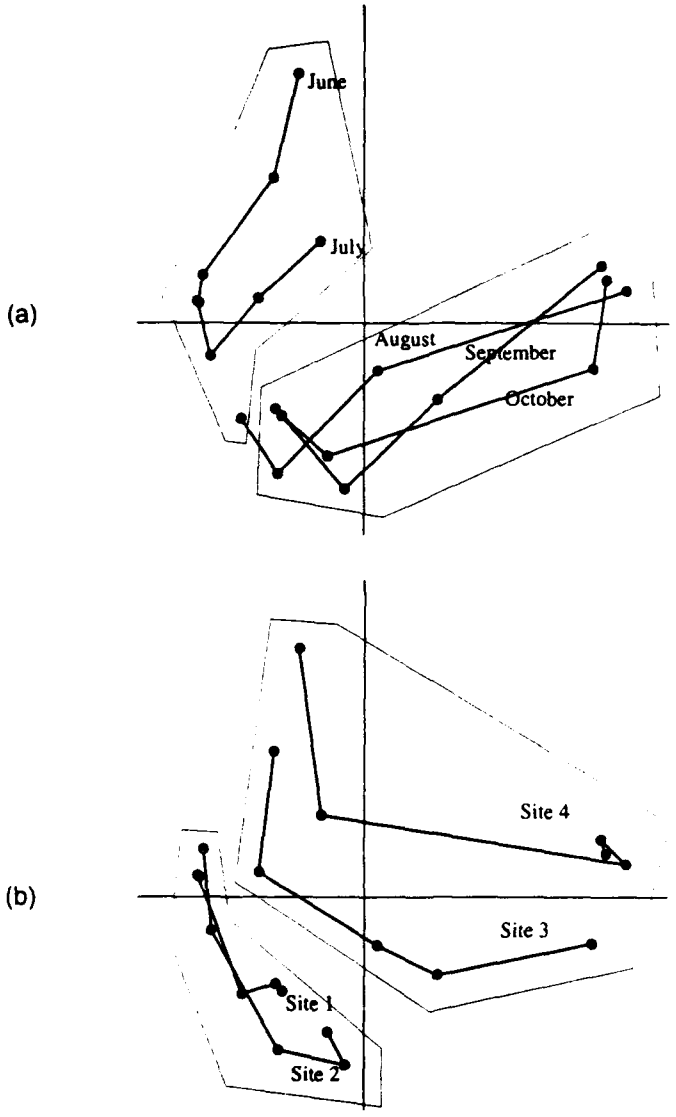


FIG. 2. Samples plotted with respect to the first two components of correspondence analysis. In Fig. 2a, heavy lines connect samples from a single date; in Fig. 2b, heavy lines connect samples from a single site. The "northwest–southeast" trend in dates and the "northeast–southwest" trend in sites are illustrated. Grouping of samples in Fig. 2a is by *k*-means clustering with squared Euclidean distance; grouping of samples in Fig. 2b is by nonmetric clustering.

classified. For our nonmetric clustering-derived characterization, there were no erroneous classifications. By comparison, we also performed "leave-one-out" testing using a linear discriminant procedure to reclassify the left-out sample. The linear discriminant misclassified 15% (3 out of 20), and this was in spite of the linear discriminant being an "informed" procedure, i.e. input to the linear discriminant procedure consisted of both the data points and an identification of which data points came from upstream samples and which from downstream samples. Nonmetric clustering is, in contrast, an "uninformed" procedure. Input to the nonmetric clustering procedure consisted only of the data points, and no information about the location of the samples. Nonmetric clustering was able to deduce the locations of the samples from the macroinvertebrate densities alone.

Discussion

Our statistical analyses supported our initial hypothesis that there were longitudinal and seasonal trends evident in the macroinvertebrate data. Ordination of samples by correspondence analysis was clearly possible (Fig. 2); however, a two-dimensional ordination was necessary to confirm each of the one-dimensional trends.

The existence of (at least) two gradients in a data set made interpretation of the data by clustering more difficult. Our two clustering techniques yielded radically different clusters because the structure of the data was complex enough to warrant two interpretations. Which trend is the "strongest" depends on how "strongest" is interpreted. In our professional judgement, the most obvious trend was the longitudinal trend. There were marked differences in the makeup of the macroinvertebrate communities from upstream and those from downstream. However, the existence of this "obvious" trend was not confirmed by *k*-means clustering. Instead, a rather new tool, nonmetric clustering, that approaches data clustering from radically different assumptions was required to "confirm the obvious."

The fact that correspondence analysis gave high scores to rare taxa might be expected because, if a taxon is rare, and only shows up at one site or date, it will, of course, be highly correlated with that site or data. But many factors can affect the reported densities of rare taxa, including drift and emergence as well as sampling technique, sorting, and taxonomic experience. Because only some of these factors are associated with a gradient, correspondence analysis may not be robust in data sets where there are many uncommon taxa. In Padden Creek, 43 of the 63 taxa were uncommon, i.e. making up less than 0.5% of the total density. The conclusion we draw is that taxa scores from correspondence analysis should not be viewed individually or in small subsets (such as the top 20), but only collectively.

Nonmetric clustering was the only technique that proved successful in both (a) confirming an observed trend and (b) providing a set of indicator taxa for that trend. Nonmetric clustering identified a subset of 15 common taxa, given in Table 2, that provided enough information to classify the samples, and did so more accurately than a linear discriminant.

Conclusion

Ecologically the dominant trends in our stream data were the longitudinal trend, where, typically, mayflies, stoneflies, and caddisflies were found at the upstream sites (Sites 1 and 2), while noninsects and tolerant taxa were found at the downstream sites (Sites 3 and 4), and the seasonal trend. Our subjective judgement was that the longitudinal trend was more significant in this study than the seasonal one. Correspondence analysis ordination of the macroinvertebrate data from Padden Creek confirmed the presence of both the longitudinal and seasonal trends in the taxa, but only as a "mixture" of each of the first two components of the ordination. In addition, correspondence analysis typically gave rare taxa the highest taxa scores, even though their relevance to large-scale trends in the data set was minor. *K*-means clustering favored the seasonal trend over the longitudinal trend, while nonmetric clustering favored the longitudinal trend. The nonmetric clustering also provided a robust means of simplifying the description of upstream and downstream clusters by identifying a set of 15 of the most com-

mon taxa that could be used to ordinate samples in other studies if a reduced sampling effort was desirable. This set of 15 proved to be a robust indicator of the location of the sample, regardless of the season in which the sample was collected. Nonmetric clustering has not previously been used to analyze benthic macroinvertebrate data, but should prove to be a useful tool for future studies, with broad applications and major advantages over current techniques.

References

ANDERSON, N. H. 1976. The distribution and biology of the Oregon Trichoptera. Technical Bulletin 134, Agricultural Experimental Station, Oregon State University, Corvallis, OR.

CULP, J. M., AND R. W. DAVIES. 1980. Reciprocal averaging and polar ordination as techniques for analyzing macroinvertebrate communities. Can. J. Fish. Aquat. Sci. 37: 1358–1365.

EDMUNDS, D. F., AND D. L. JENSEN. 1976. The mayflies of North and Central America. University of Minnesota Press, Minneapolis, MN.

FAITH, D. P., AND R. H. NORRIS. 1989. Correlation of environmental variables with patterns of distribution and abundance of command and rare freshwater macroinvertebrates. Biol. Conserv. 50: 77–98.

FISHER, D., AND P. LANGLEY. 1986. Conceptual clustering and its relation to numerical taxonomy, p. 77–116. *In* W. A. Gale [ed.] Artificial intelligence and statistics. Addison-Wesley, Reading, MA.

FISHER, S. G. 1983. Succession in streams, p. 7–27. *In* J. R. Barnes and G. W. Minshall [ed.] Stream ecology, application and testing of general ecological theory. Plenum Press, New York, NY.

GAUCH, H. G., R. H. WHITTAKER, AND T. R. WENTWORTH. 1977. A comparative study of reciprocal averaging and other ordination techniques. J. Ecol. 65: 157–174.

GREEN, R. H. 1974. Multivariate niche analysis with temporally varying environmental factors. Ecology 55: 73–83.

GOODMAN, L. A., AND W. H. KRUSKAL. 1954. Measures of association for cross classifications. Am. Stat. Assoc. J. 49: 732–764.

HACHMOLLER, B. 1989. The benthic macroinvertebrate community in Padden Creek, summer, 1988. M.S. thesis, Western Washington University, Bellingham, WA.

HACHMOLLER, B., R. MATTHEWS, AND D. BRAKKE. 1990. Effects of riparian community structure, sediment size, and water quality in a small, suburban stream. Northwest Sci. 65: 125–132.

HATCH, M. H. 1953–65. The beetles of the Pacific Northwest. Vol. I–IV. University of Washington Press, Seattle, WA.

HYNES, H. B. N. 1970. The ecology of running waters. University of Toronto Press, Toronto, Ont.

JAIN, A. K., AND R. C. DUBES. 1988. Algorithms for clustering data. Prentice Hall, Englewood Cliffs, NJ.

JEWETT, S. G. 1959. The stoneflies (Plecoptera) of the Pacific Northwest. Oregon State College Press, Corvallis, OR.

KENKEL, N. C., AND L. ORLOCI. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 67: 919–928.

LUDWIG, J. A., AND J. F. REYNOLDS. 1988. Statistical ecology. John Wiley, New York, NY.

MATTHEWS, G. B., AND J. W. HEARNE. 1991. Clustering without a metric. IEEE Trans. Pattern Anal. Mach. Intell. 13: 175–184.

MATTHEWS, R. A., G. B. MATTHEWS, AND W. J. EHINGER. 1991. Classification and ordination of limnological data: a comparison of analytical tools. Ecol. Model. 53: 167–187.

MCCAFFERTY, W. P. 1981. Aquatic entomology. Jones and Bartlett Publishers, Inc., Boston, MA.

MERRITT, R. W., AND K. W. CUMMINS. 1984. An introduction to the aquatic insects of North America. 2nd ed. Kendall/Hunt Publishing Company, Dubuque, IA.

MINCHIN, P. R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69: 89–107.

MINSHALL, G. W. 1988. Stream ecosystem theory: a global perspective. J. N. Am. Benthol. Soc. 7: 263–288.

PENNAK, R. W. 1978. Freshwater invertebrates of the United States. John Wiley and Sons, New York, NY.

RICKER, W. E., AND G. G. E. SCUDDER. 1975. An annotated checklist of the Plecoptera (Insecta) of British Columbia. Syesis 8: 333–348.

ROSS, H. H. 1937. Nearctic alder flies of the genus *Sialis*. Ill Nat. Hist. Surv. Bull. 21: 57–78.

SCHAEFFER, D. J., AND J. A. PERRY. 1986. Gradients in the distribution of riverine benthos. Freshwater Biol. 16: 745–757.

SHELDON, A. L., AND R. A. HAICK. 1981. Habitat selection and association of stream insects: a multivariate analysis. Freshwater Biol. 11: 395–403.

STARK, B. P., AND A. R. GAUFIN. 1976. The nearctic genera of Perlidae. Misc. Publ. Entomol. Soc. Am. 10(1).

STONE, A., C. W. SABROSKY, W. W. WIRTH, R. H. FOOTE, AND J. R. COULSON. 1965. A catalog of the Diptera of America north of Mexico. Agricultural Research Service, United States Department of Agriculture, Washington, D.C.

TER BRAAK, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67: 1167–1179.

UHLIG, L. J. 1991. Comparison of predator-prey relationships between stoneflies and mayflies in various habitats of Padden Creek. M.S. thesis, Western Washington University, Bellingham, WA.

Appendix: Nonmetric Clustering

We give here a brief introduction to the technique of nonmetric clustering, which is described fully in (Matthews and Hearne 1991). Traditional clustering algorithms, such as *k*-means clustering, rely on a metric, or distance measure, defined over *n*-dimensional space. Points are then divided into clusters based on cluster "quality," where quality is in turn based on simultaneously minimizing intracluster distance and maximizing intercluster distance. In Fig. A.1, for instance, the points in the upper right would constitute one cluster because they are all close to each other, and the points in the lower left would constitute the second cluster because they are all close to each other and at the same time far from the points in the other cluster.

Problems arise with this method when other dimensions are added, however. In Fig. A.2a, the points all have the same x and y coordinates as in the previous figure, but a random value for the z dimension has been added. Intuitively, the points are still in the same clusters, and the third dimension represents pure noise that should be ignored. Metric-based clustering, however, must compose a metric out of all dimensions, with the result that the clusters proposed for the data are as shown in Fig. A.2b. If metric-based clustering is to succeed at all, some kind of data transformations or weighted metrics must be employed.

Nonmetric clustering, on the other hand, is not based on an *n*-dimensional metric. Instead, each dimension is examined independently of the others, and the association between the clustering and the dimension is measured. In Fig. A.1 the association between the obvious clusters and each of the x and y axes is evident. A quantitative measurement of this association is used to indicate the strength of the association. Guttman's λ (Goodman and Kruskal 1954), which is similar to a chi-squared statistic, is used for reasons discussed by Matthews and Hearne (1991). The optimal clustering, then, is selected as the one that has the strongest association with the largest number of dimensions. The dimensions themselves are not combined into a metric, and there is no call to include all dimensions in the estimate of clustering quality.

For our example data set, the nonmetric clustering for three dimensions, shown in Fig. A.2c is identical to the obviously "correct" clustering in two dimensions. This is because the best associations between clustering and dimensions are with the x and y axes. There is no way a clustering can be found that will associate well with more than two axes, and so only the x and y axes are used to measure clustering quality. The z axis is ignored.

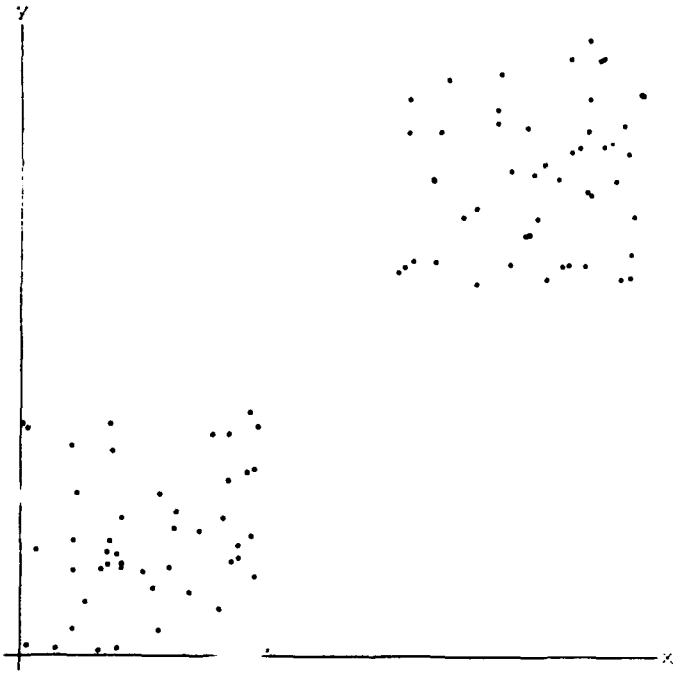


FIG. A.1. Artificially generated data set clustered in two dimensions.

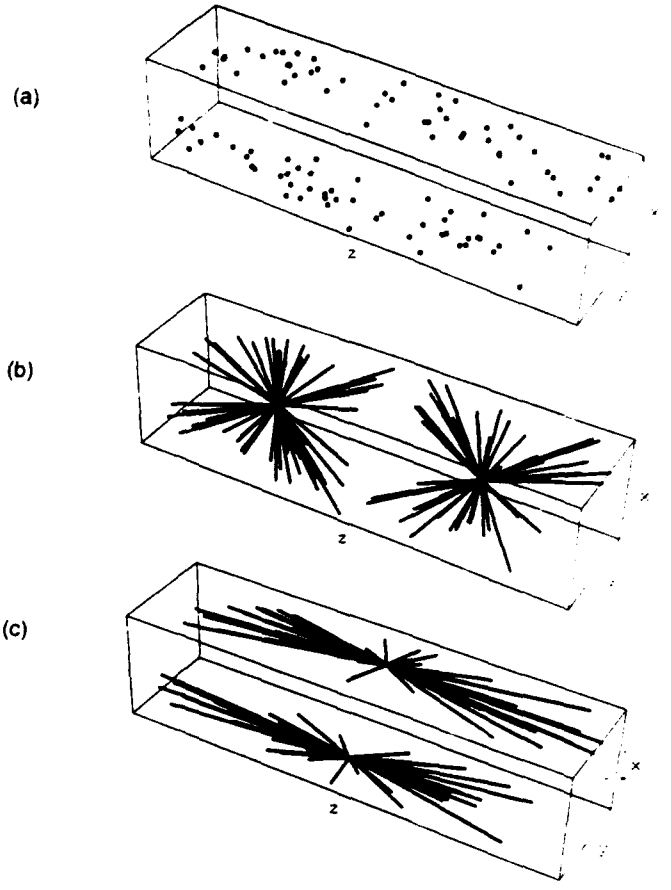


FIG. A.2. Artificial data set of Fig. A1 with (a) a random component in the z dimension added, (b) clustering by k-means, and (c) nonmetric clustering.

A computer program, called RIFFLE, implementing nonmetric clustering, has been constructed and is described in Matthews and Hearne (1991). RIFFLE was used for all nonmetric clustering discussed in this paper.

Nonmetric clustering thus offers the following advantages over traditional methods: (1) it does not combine counts from dissimilar taxa by means of sums of squares, or other ad hoc mathematical techniques; (2) it does not require transformations of the data, such as normalizing the variance; (3) it works without modification on incomplete data sets; (4) it can work without further assumptions on different data types (e.g. species counts or presence/absence data); (5) significance of a taxon to the analysis is not dependent on the absolute size of its count, so that taxa having a small total variance, such as rare taxa, can compete in importance with common taxa, and taxa with a large, random variance will not automatically be selected, to the exclusion of others; (6) it provides an integral measure of "how good" the clustering is, i.e. whether the data set differs from a random collection of points; and (7) it can, in some cases, identify a subset of the taxa that serve as reliable indicators of the physical environment; in our case, the indicator species were proved, in testing, to be more reliable than indicators based on a linear discriminant.

The primary disadvantages of nonmetric clustering, as we see them, are as follows. (1) There are some cases, documented in (Matthews and Hearne 1991), where metric clustering is to be preferred over nonmetric clustering. In general, we recommend using both, and examining the results critically, rather than accepting a single clustering method as the best for all cases. (2) The RIFFLE implementation of nonmetric clustering is very computer intensive, and takes much longer to run than k-means clustering. (3) Implementations of the technique, such as RIFFLE, are not widely available yet.

Ecological Modelling, 53 (1991) 167–187
Elsevier Science Publishers B.V., Amsterdam

Classification and ordination of limnological data: a comparison of analytical tools

R.A. Matthews [a], G.B. Matthews [b] and W.J. Ehinger [c]

[a] *Huxley College of Environmental Studies,*
[b] *Computer Science Department,*
Western Washington University, Bellingham, WA 98225, U.S.A.
[c] *Department of Biology, University of North Carolina, Chapel Hill, NC 27514, U.S.A.*

(Accepted 7 May 1990)

ABSTRACT

Matthews, R.A., Matthews, G.B. and Ehinger, W.J., 1991. Classification and ordination of limnological data: a comparison of analytical tools. *Ecol. Modelling*, 53: 167–187.

In this paper we compare the differences between principal components analysis, hierarchical clustering, correspondence analysis and conceptual clustering to show their effectiveness for identifying patterns in a large limnological data set. The data for this comparison come from a multi-year study of Lake Whatcom, a large lake located in the Puget Sound lowlands of the state of Washington. The data include both physical and chemical parameters (temperature, dissolved oxygen, pH, alkalinity, turbidity, conductivity, and nutrients) as well as biological parameters (Secchi depth, chlorophyll *a*, and phytoplankton species and total counts). The patterns we expected to find include (a) temperature and dissolved oxygen interactions, (b) ordination by algal bloom sequences, and (c) clustering due to the effects of stratification.

Principal components analysis was somewhat useful for confirming known water quality trends, but did not successfully identify large-scale patterns such as stratification and seasonal plankton changes. Correspondence analysis proved to be superior to principal components analysis for detecting phytoplankton trends, but was not as good for interpreting water quality changes. Hierarchical clustering produced highly unbalanced trees for both the water quality and phytoplankton data, and was useless as an exploratory tool. A new approach to clustering, implemented in the computer program RIFFLE, is introduced here. This clustering algorithm outperformed the other exploratory tools in clustering and parameter ordination, and successfully identified a number of expected and unexpected patterns in the limnological data.

INTRODUCTION

One of the most difficult problems in aquatic ecology is the interpretation and modelling of the complex data sets that are generated from limnological

0304-3800/91/$03.50

research. The data generally are not linear, rarely conform to parametric assumptions, and are often measured using incommensurable units such as length, concentration, and frequency. In addition, most limnological research generates incomplete data sets, not only because of sample loss, but also due to sampling design. For example, lake depth, temperature, and dissolved oxygen may be measured every few meters from the surface to the bottom, while plankton populations are usually sampled only in the photic zone. As a result, we may have to rely on the robustness of a statistical test to identify significant trends despite violation of the test's fundamental assumptions. Further, true gradients, as understood in terrestrial ecology, are rarely present. Nevertheless, patterns of algal blooms and successions are present, and their recognition poses an important problem for data analysis and modelling.

In this paper we compare several types of analytical procedures, including graphical analysis, hierarchical clustering, and ordination (principal components analysis and correspondence analysis), to see how well they identify patterns in a large limnological data set. While all of these methods are in common use, they are not all equally useful for identifying patterns in ecological data sets (Pielou, 1984; Ludwig and Reynolds, 1988). In addition, we used a new version of conceptual clustering (Fisher and Langley, 1986), which turned out to be markedly superior to correspondence analysis in parameter ordination, and superior to hierarchical techniques in clustering.

Our data come from Lake Whatcom, a large monomictic lake in Washington. Water quality data have been collected from Lake Whatcom since the early 1960's, with intensive sampling since 1982. The data for this paper are from spring 1987 through winter 1988 because this period included intensive plankton sampling as well as water quality monitoring. The patterns we expected to find in the lake included: (a) temperature and dissolved oxygen interactions, (b) algal bloom sequences, and (c) indicators and effects of stratification. Evidence for all of these was discovered in the data set. However, some of the analytical techniques were less useful than others for identifying the limnological trends. We have included a general discussion of the fundamental differences between each analytical technique as well as a summary of the strengths and weaknesses of each technique for identifying patterns in limnological data.

METHODS

Study site

Lake Whatcom is a 2000 ha chain lake located in the Puget Sound lowlands of northwestern Washington (Fig. 1). The lake is divided into three

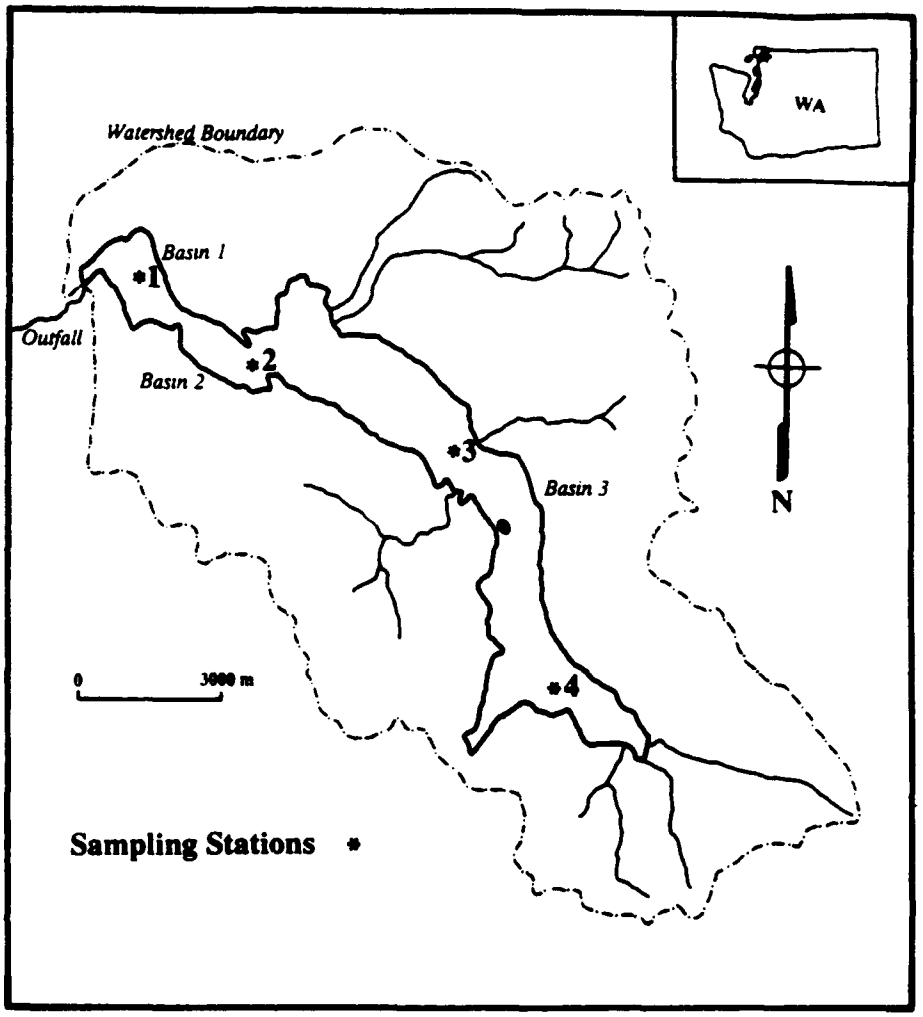


Fig. 1. Lake Whatcom sampling sites, Whatcom County, Washington.

distinct basins by subsurface sills; the largest basin, Basin 3, contains 96% of the lake volume, while Basins 1 and 2 each contain about 2% of the total lake volume (Lighthart et al., 1972). Lake Whatcom is a warm, monomictic lake; the direction of flow is from Basin 3 → Basin 2 → Basin 1. All of the perennial streams in the Lake Whatcom watershed drain into Basin 3. The only natural outflow from the lake is Whatcom Creek in Basin 1. However, the city of Bellingham withdraws water from Basin 2 for municipal drinking water and industrial uses. In the summertime the municipal withdrawal is often the only significant outflow from the lake.

Water quality and phytoplankton sampling

Water samples were collected at four sites in Lake Whatcom (Fig. 1) from March 1987 to October 1988. Temperature, pH, conductivity, and dissolved oxygen were measured in the field using a Hydrolab Surveyor II. In Basins 1 and 2, where the maximum depths are 20 and 22 m, respectively, these measurements were taken at 2-m intervals from the surface to the bottom of the water column. In Basin 3 (maximum depth > 90 m), the measurements were taken at 2-m intervals to the depth of 20 m, and at 5-m intervals from 20 m to the bottom. Secchi depth was also measured in the field at each site.

The water samples for nutrients analyses (ammonia, nitrate/nitrite, total nitrogen, soluble reactive phosphate, and total phosphorus), total organic carbon, and dissolved inorganic carbon analyses were collected at 5-m intervals in Basins 1 and 2, and 10-m intervals in Basin 3. The nutrient analyses were done using a Technicon Autoanalyzer, following EPA (1983) guidelines for sampling handling and analysis. The total organic carbon and dissolved inorganic carbon analyses were done using an OIC Model 0524B Infrared Carbon Analyzer (APHA, 1985).

All chlorophyll and phytoplankton samples were collected at 5-m intervals from the surface to 15 m (phytoplankton) or 20 m (chlorophyll). Chlorophyll *a* extractions were done by filtering 250–500 mL of sample through a glass fiber filter, which was ground in a tissue grinder and extracted with 90% spectrophotometric grade acetone. The chlorophyll *a* concentrations, corrected for phaeophytin *a*, were measured using a calibrated Turner Designs fluorometer (APHA, 1985). Phytoplankton samples were preserved with Lugol's solution, and were identified and counted using a Sedgewick-Rafter counting chamber on an Olympus Inverted Microscope (APHA, 1985; Lind, 1985). Representative phytoplankton samples were sent to the Academy of Natural Sciences of Philadelphia for taxonomic verification.

Data analysis methods

The data were analyzed using either ordination, clustering, or both. Ordination of 'points' (all measurements collected at a particular date, site, and depth, sometimes called 'samples' or 'sampling units') was done by principal components and correspondence analysis (reciprocal averaging). Ordination of 'parameters' (e.g., pH, temperature, etc., sometimes called 'attributes", 'dimensions', or 'variables') was done by correspondence analysis and conceptual clustering. Clustering was done with an agglomerative, hierarchical algorithm, as well as with an optimizing, conceptual clustering

algorithm. Visual confirmation of patterns in the data was made using two- and three-dimensional graphical displays of the data.

Point ordination

Principal components analysis was done using data normalized by mean and standard deviation (z-scores), using the FACTOR procedure provided in the SPSS-X statistical package. This resulted in several ordinations of the points, one for each principal component. Generally, the first three or four principal components were inspected graphically.

Correspondence analysis (reciprocal averaging), which simultaneously ordinates both the parameters and the data points, has proven better than principal components analysis in the analysis of many kinds of ecological data. In data sets involving large-scale gradients in the environment, for example, with high beta diversity along the gradients, correspondence analysis outperforms principal components analysis (Kenkel and Orloci, 1986). It can be used for detecting unknown gradients or confirming the existence of expected ones. Correspondence analysis scores were computed directly using the iterative technique (Pielou, 1984, pp. 184–188).

Hierarchical clustering

Hierarchical clustering uses a measure of similarity or distance between points, and derived measures of inter-cluster and intra-cluster distance. It is hierarchical in that each cluster is a subcluster of a larger cluster; the total clustering forms a tree, or dendrogram. Balanced dendrograms indicate a good clustering into roughly equal-sized clusters, while unbalanced dendrograms indicate little real clustering, but instead a gradual agglomeration of sample points into a single group.

The choice of a distance measure is often critical to hierarchical clustering (Ludwig and Reynolds, 1988). We employed two distance measures for hierarchical clustering: squared Euclidean distance, defined as $\sum_i (x_i - y_i)^2$ and cosine of vectors distance, defined as $\sum_i (x_i y_i) / \sqrt{(\sum_i x_i^2)(\sum_i y_i^2)}$, where x_i and y_i are the parameter values for two points. Cosine distance is similar to chord distance (Ludwig and Reynolds, 1988), and considers only the relative proportions of the various parameters that make up a sample point. Squared Euclidean distance also takes into account the absolute size of parameter values.

The algorithm we used for forming the hierarchy of clusters was average linkage between clusters. This method gives good results on synthetic, Gaussian data known to have well-defined clusters (Bayne et al., 1980).

Conceptual clustering

The philosophical difficulty with hierarchical clustering is that it assumes the meaningfulness of combinations of parameters, such as the Euclidean and cosine distances, above. In ecological data sets, such compositions as these two are often not meaningful, due to incommensurability. For example, an uncommon organism with a large individual biovolume may have the same total biomass as a common organism with a smaller individual biovolume, but since both species are measured in organisms per L, the common organism will dominate in terms of absolute number and proportion. Predators, for example, often fall into this category, being generally large in size but small in number. However, their functional importance would be overlooked by this analytical technique which would simply add or multiply the two numbers. The problem lies not in the manner of counting organisms, but in the necessity to *combine* counts of dissimilar species. The problem is even worse for water quality data, where different parameters are measured in degrees, pH units, concentrations, and so on.

Conceptual clustering can be used as an alternative to hierarchical clustering [see Fisher and Langley (1986) for a survey]. A clustering technique is called 'conceptual' if it yields descriptions of the clusters in terms of *concepts*, i.e., in terms of only conceptually important parameters. What is 'conceptually important' depends on context, but in scientific data analysis we take the following as an acting principle: Clusters are conceptually important if knowledge of such clusters increases the reliability of predictions about parameter values. In other words, we seek clusters such that most (if not all) of the actual observed data values for a sample can be predicted more accurately after its cluster has been identified than before such identification. Thus, 'conceptually important' clusters, in our methodology, are those that warrant accurate predictions of parameter values.

We developed a clustering tool, called RIFFLE, in line with these principles, which is superior to traditional clustering methods for a wide range of ecological data sets (Matthews and Hearne, 1991). A brief description of the algorithm is given in Appendix A. RIFFLE has the following advantages over traditional clustering methods: (1) Measures based on combinations of incommensurable parameters, such as Euclidean distance in parameter space are not used, (2) transformations of scale do not affect the outcome, (3) parameters can be nominal, ordered, numeric, or mixed, (4) 'noisy' parameters, i.e., those with large variance but little association with any other parameters, are automatically filtered out and have little effect on the resulting clustering, (5) 'rare' parameters, i.e., those with small variance but with a significant correlation to the dominant patterns of the data set, are automatically given weight in accord with that correlation, and (6) no

assumptions about points with missing values, such as replacement with zeroes or with the mean, need to be made. RIFFLE simultaneously clusters the data and ordinates the parameters in terms of their conceptual significance to the clusters. It is thus, in a sense, similar to correspondence analysis in that simultaneous analysis of points and parameters is done, except that a *non-linear* patterning the points (a clustering) is sought together with a *linear* ordination of the parameters. Correspondence analysis attempts to provide a linear ordination of both.

RESULTS AND DISCUSSION

Physical-chemical data

The physical-chemical data from Lake Whatcom indicate that the three basins are dissimilar, which is best illustrated by comparing graphs of the temperature and dissolved oxygen data for the four sites (Figs. 2–5). The two shallow basins (Basin 1, Site 1 and Basin 2, Site 2) both had significant oxygen deficits, and both developed anoxic hypolimnia during the summer. Basin 3, Site 3, experienced some oxygen depletion during the summer; however, the oxygen concentrations usually did not fall below 2 mg/L. Basin 3, Site 4 maintained consistently high dissolved oxygen levels throughout summer stratification, even at the bottom of the water column.

The oxygen deficit in Basin 1 was more pronounced than in Basin 2. This observation was discussed by Ehinger (1988) and is thought to be due, at least in part, to isolation of Basin 1 during the summer when the outflow from the lake into Whatcom Creek was reduced to near zero. The City of Bellingham continued to withdraw water from Basin 2 throughout the summer, which flushed Basin 2 with high quality water from Basin 3.

The remaining water quality parameters were strongly influenced by the temperature and dissolved oxygen conditions in the lake. Basins 1 and 2 experienced epilimnetic nitrate depletion during summer algal blooms. Concurrently, ammonia and phosphate were released from the sediments and accumulated in the hypolimnia of both basins. In Basin 3, similar conditions developed, but to a much lesser extent. Alkalinity and pH values showed little variation except during stratification. During this time, the pH values were slightly higher in the epilimnia of Basins 1 and 2 due to photosynthetic activity, while the pH values in the hypolimnia were lower due to the release of reduced compounds from the sediments. Similarly, the alkalinity values increased slightly near the sediments during stratification. Conductivity, turbidity, dissolved inorganic carbon, and total organic carbon values were fairly uniform throughout the sampling period. A complete listing of the

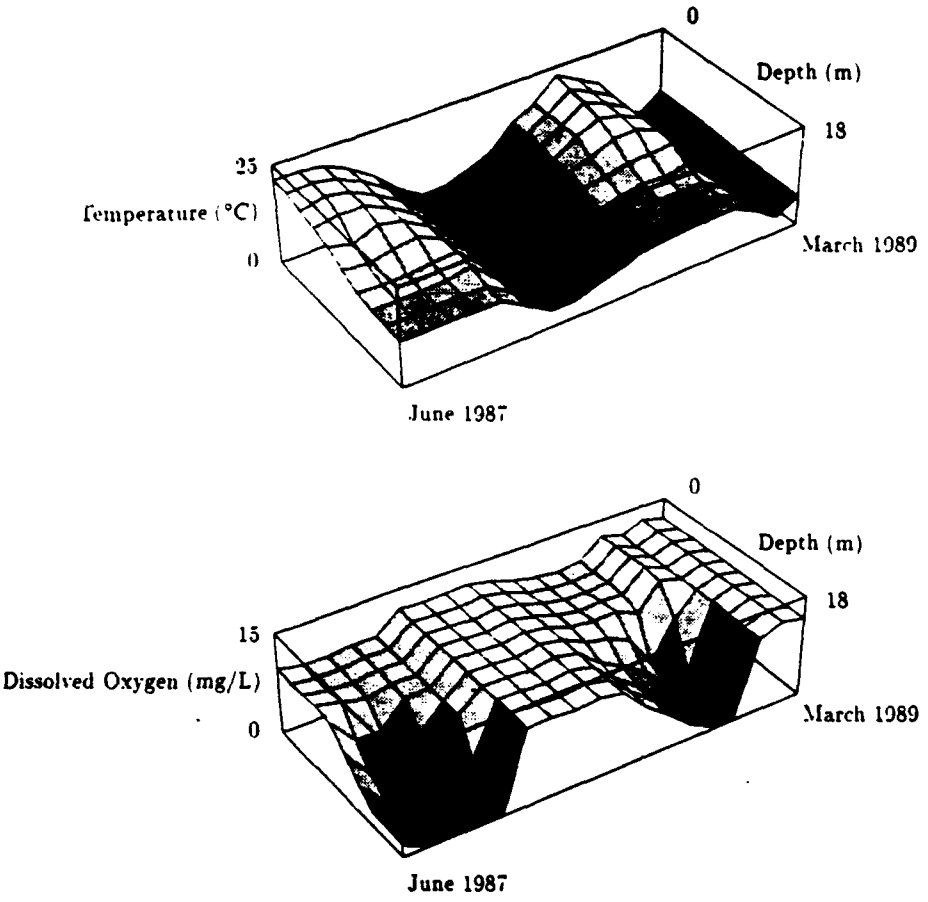


Fig. 2. Temperature and dissolved oxygen profiles for Basin 1, Site 1.

water quality data is available from the authors, and a list of parameters sampled is in Appendix B.

Conceptual clustering of the physical-chemical data proved to be best at confirming the expected trends. Figure 6 shows how RIFFLE clustered the physical and chemical data for each discrete sample set (matched by date, site, and depth class). The RIFFLE clusters were plotted by the date and temperature value for the data set so that the influences of thermal stratification can be observed. Sample points were grouped into classes based on approximate ($\approx$ 5 meter) depth, and data values were taken as averages of the values in a single depth class. Depth classes were used because of the large number of points in the Hydrolab data sets (> 1600 for each parameter) and because there was some variation in the depth of some samples. For example, the 'bottom' measurements varied by several meters, depending on

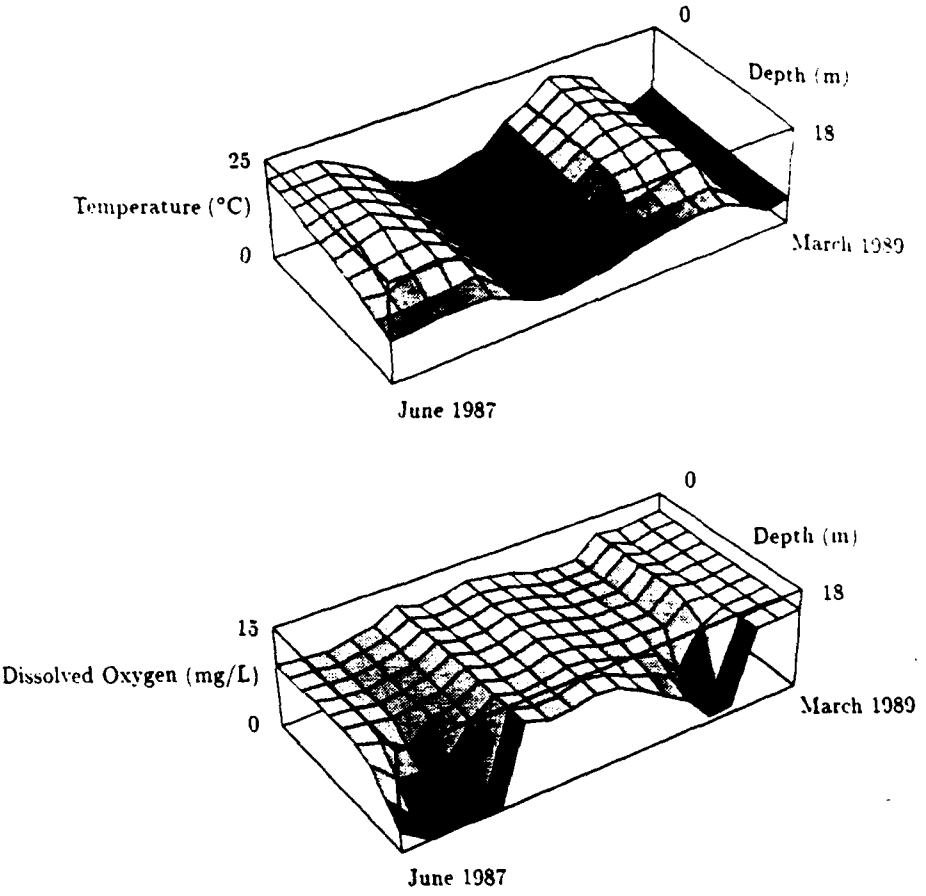


Fig. 3. Temperature and dissolved oxygen profiles for basin 2, Site 2.

where the boat was located. A smaller total number of points also helped in the graphical presentation of the data.

In Basin 1, three clusters were selected as best describing the data. Two of the clusters (○ and ▷) separate the epilimnion and hypolimnion samples during stratification, while the third cluster (★) identifies the well-mixed samples of the unstratified period. The vertical lines marking stratification and turnover were estimated from the temperature data for each basin; however, the exact timing of these events was not determined. This is important because most of the misclassifications in the RIFFLE clusters occurred within one sampling date of our estimated dates for stratification or turnover.

Basin 3 clustered into only two groups: stratified epilimnial samples (○) and a second group consisting of both hypolimnial samples and mixed lake

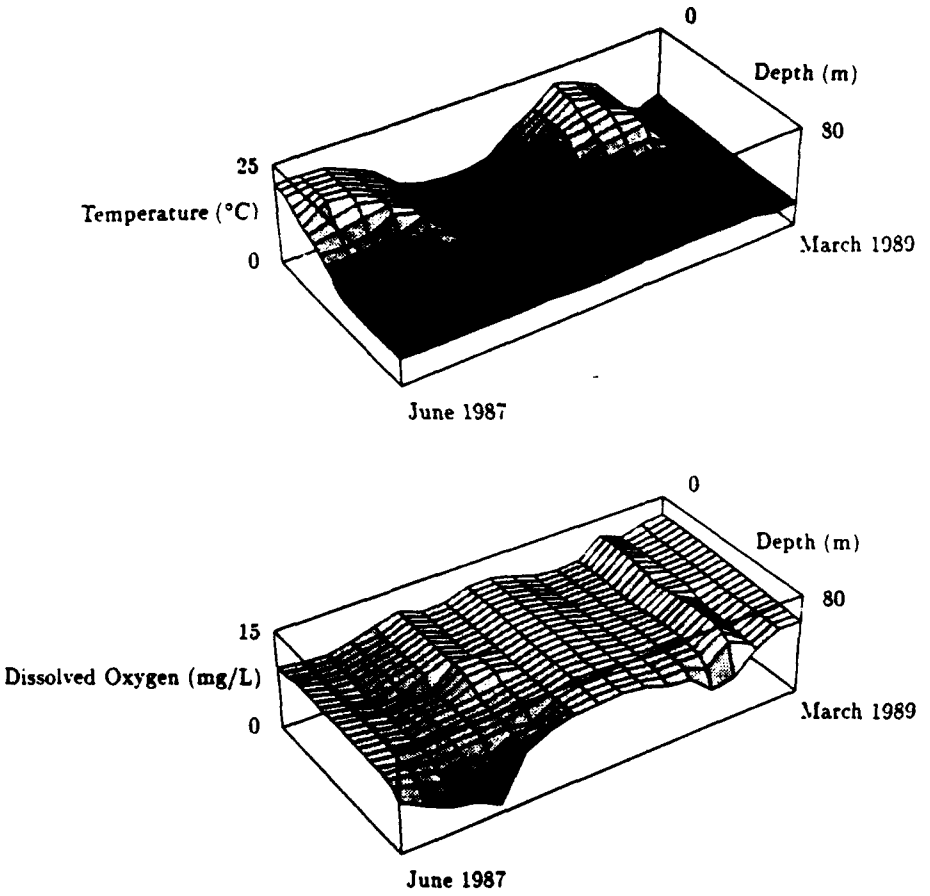


Fig. 4. Temperature and dissolved oxygen profiles for Basin 3, Site 3.

samples (★). This supports our temperature and dissolved oxygen data that show Basin 3 to be oligotrophic, with little change in the hypolimnetic water quality occurring during summer stratification.

In Basin 2, a unexpected pattern emerged. During stratified periods, three clusters were identified. Upon closer inspection of the temperature and dissolved oxygen data, we found that the depth of the thermocline was deeper in Basin 2 than in Basin 1, and the height of the anoxic portion of the hypolimnion (0–2 mg/L) was much higher in Basin 1 than in Basin 2. In Basin 1, both the surface and the 10-m depth classes would lie primarily in the epilimnion, while the remaining measurements (20 m and bottom) would be in the hypolimnion, and strongly influenced by anoxic conditions. However, in Basin 2, the 10-m depth class would be at the thermocline and slightly above the anoxic portion of the hypolimnion. The remaining sam-

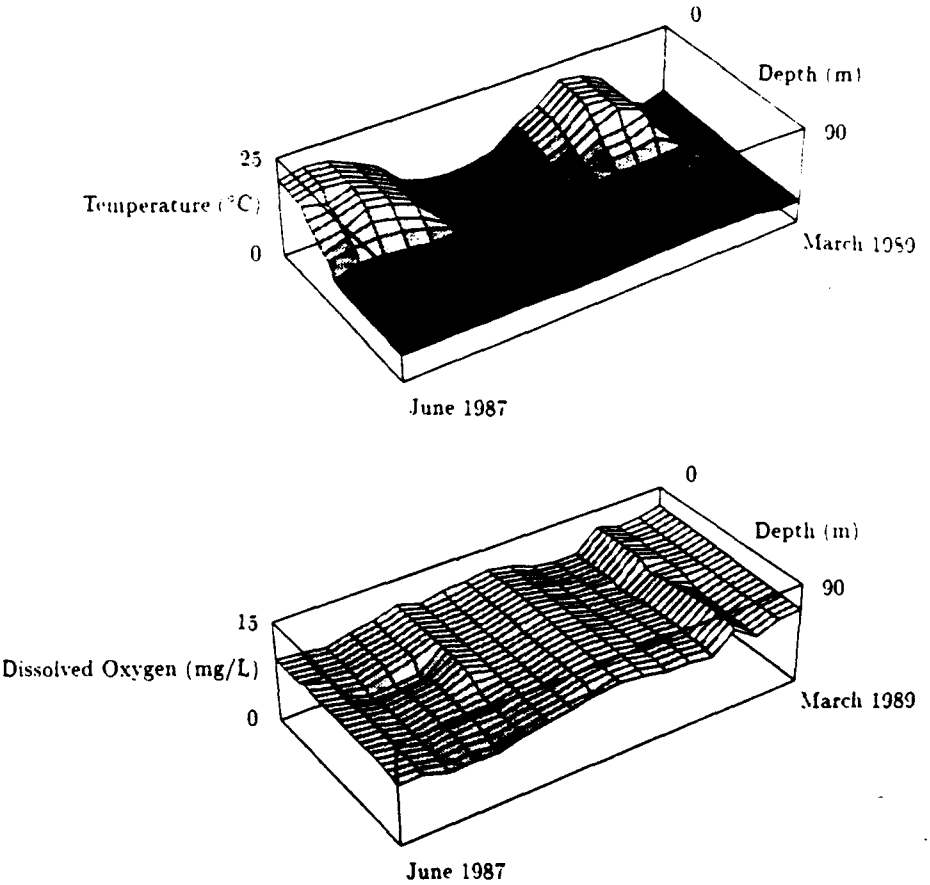


Fig. 5. Temperature and dissolved oxygen profiles for Basin 3, Site 4.

ples (at 20 m and below) would reflect hypolimnetic influences. The three clusters in Basin 2, therefore, identify the epilimnion, metalimnion, and hypolimnion.

Principal components analysis did not work well when plotted by individual basins, but did identify the major trends for the entire data set: The first principal component accounted for 24% of the total variance; its dominant terms (with a factor greater than 0.5) were:

$$0.872\ \text{Temperature} - 0.842\ \text{Depth} + 0.735\ \text{pH} - 0.623\ \text{Nitrate/Nitrite}$$

The second principal component accounted for another 19% of the total variance and its dominant terms were:

$$-0.779\ \text{Dissolved Oxygen} + 0.695\ \text{Turbidity} + 0.663\ \text{Alkalinity}$$

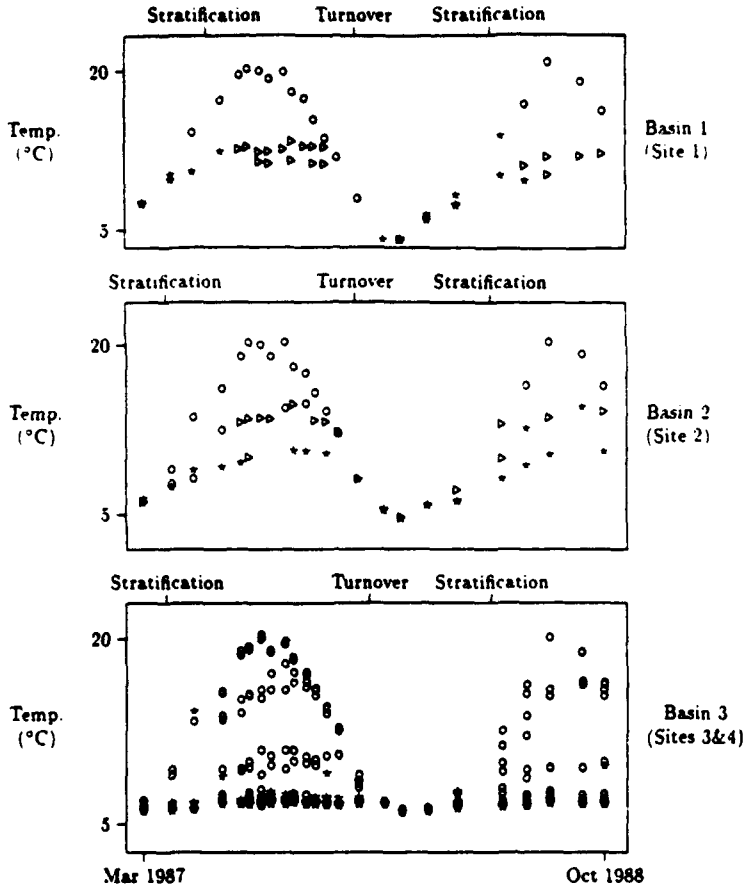


Fig. 6. RIFFLE clustering of chemical data. Conceptual clusters (○, ▷, and ★) plotted by temperature and date.

The first principal component identified the inverse relationship between temperature and depth during summer stratification as well as the changes in pH and nitrate values that were discussed earlier for Basins 1 and 2. The second component picked up on the hypolimnetic oxygen depletion that was observed, to a greater or lesser extent, in all three basins following stratification. The positive turbidity factor was probably an artifice that resulted from sampling too near the sediments, while the alkalinity factor again reflects the effects of biological activity during stratification.

Hierarchical clustering and correspondence analysis did not identify any meaningful trends in this data set. Correspondence analysis found nearly all points to have the same scores, and thus any parameter ordination was of doubtful validity. Hierarchical clustering resulted in unbalanced dendro-

grams, and had the added disadvantage that, since points with missing data could not be included, the data had to be severely subsetted. Several parameters (Secchi depth, dissolved inorganic carbon, and total organic carbon) had to be excluded because they were measured less frequently than other parameters.

Phytoplankton data set

Since it is only useful to collect phytoplankton data at or near the surface, this data set is considerably smaller, in terms of number of points, than the physical-chemical data set. A complete listing of taxa found is provided in Appendix C.

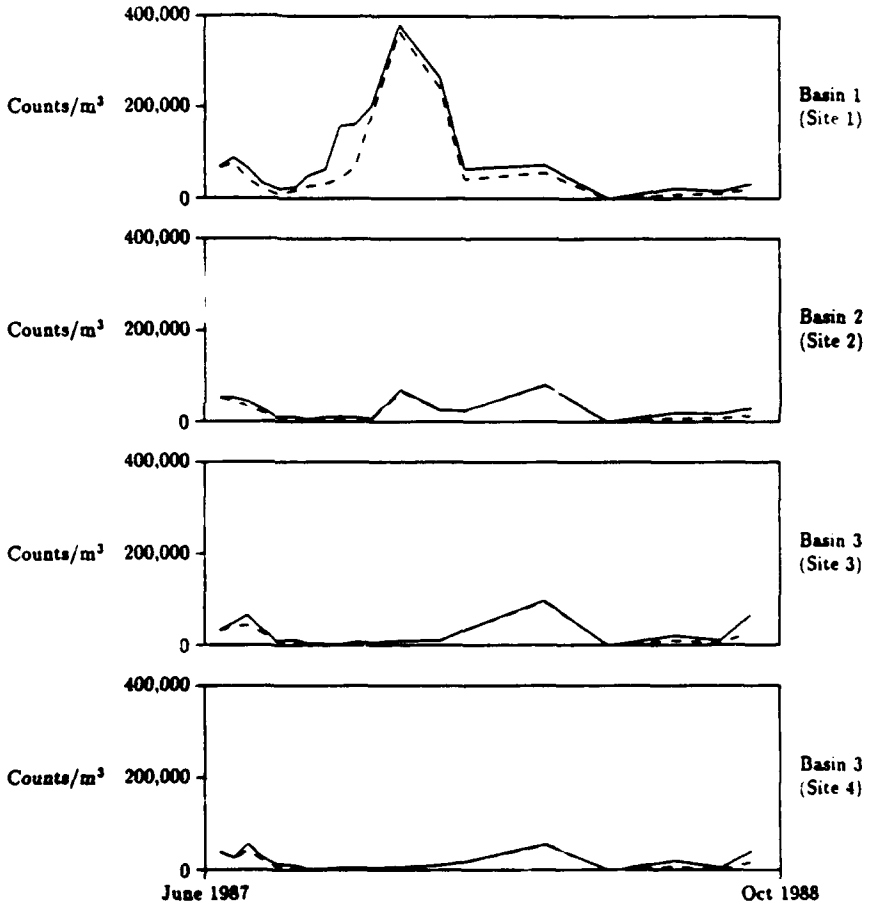


Fig. 7. Total phytoplankton (solid) and diatoms (dashed) in Lake Whatcom.

Figure 7 shows a summary of the phytoplankton data for Lake Whatcom. Diatoms (predominantly *Melosira ambigua* (Grun.) O. Mull, *Melosira distans* (Ehr.) Bethge, and *Fragilaria crotonensis* (Kitt.) dominated the phytoplankton populations most of the year, with peaks occurring during the winter and spring.

During the late summer (during periods of nutrient depletion in the epilimnion), blooms of mostly green and bluegreen algae developed, especially in Basin 1. The densities of green and bluegreen algae never reached the peak densities that were measured for the winter/spring diatom blooms. This is partly due to our system of counting, whereby *Coelosphaerium naegelianum* Unger, a common late summer bluegreen alga, was counted by colonies rather than individual cells. If *Coelosphaerium* had been counted by individual cells (not an easy task) or if each plankton count was weighted to account for biovolume, [as in Ehinger (1988)], the *Coelosphaerium* total

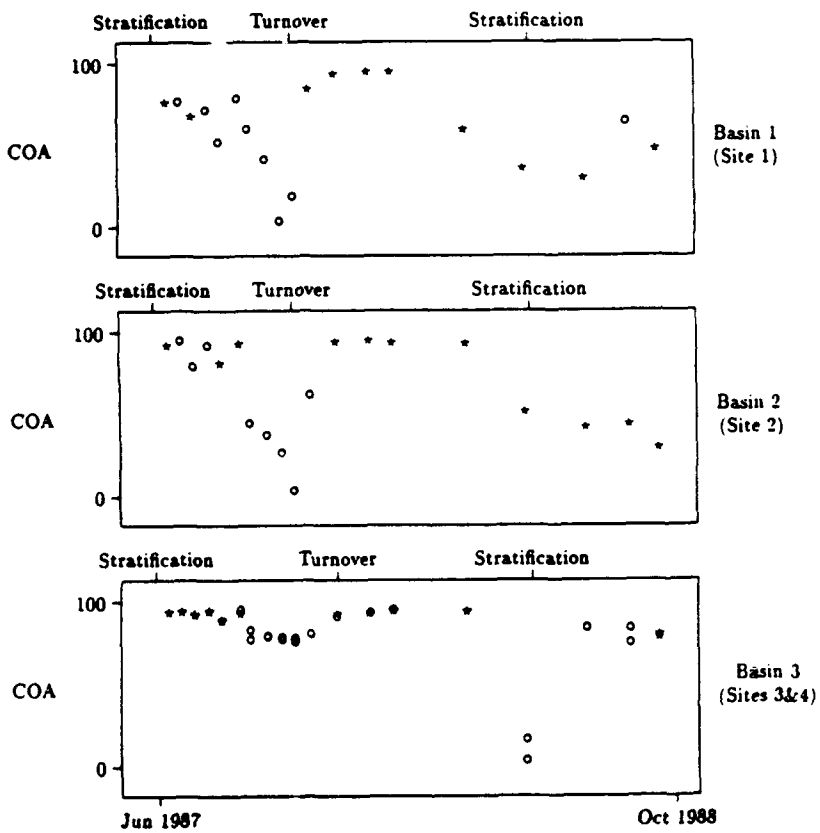


Fig. 8. RIFFLE clustering of phytoplankton data. Conceptual Clusters (○ and ★) plotted by correspondence analysis score and date.

'count' would increase. This problem of counting individuals, colonies, or biovolumes is frequently encountered in limnological data sets, and is part of the reason why the statistical tool needs to be insensitive to scale.

Conceptual clustering of the phytoplankton data again proved valuable for identifying the major trends in the lake. Figure 8 shows the clusters generated by RIFFLE plotted by correspondence analysis score (COA) vs. time. In all three basins, samples collected before and after turnover tended to be in different clusters; similar rapid changes in the phytoplankton populations did not occur following stratification. Turnover is a dramatic event in lakes, often occurring within a few days, that causes rapid changes in the water quality of the lake. Stratification, however, causes a gradual divergence of the water quality in the epilimnion and hypolimnion. In Fig. 8, late summer phytoplankton (○) were clearly distinguished from post-turnover phytoplankton (★). However, the late summer phytoplankton populations were not reestablished until several months after the onset of stratification.

In creating the clusters shown in Fig. 8, RIFFLE clustered temporally adjacent points. This is in line with the proposed existence of temporal 'plateaus' in phytoplankton succession, mentioned in (Legendre et al., 1985). RIFFLE, however, clustered them successfully without the ad hoc imposition of an explicit chronological constrait or the elimination of singleton clusters.

The taxa that contributed most heavily to the RIFFLE clusters included many common species (e.g *Fragilaria* and *Coelosphaerium*), but also included several 'rare' species that were highly correlated with turnover. One example is *Ceratium hirudinella* (O.F. Muell.), a large dinoflagellate, that never occurred in large numbers, but was only collected during late summer just prior to turnover. *Ceratium* is able to compete well during late summer because it can swim to positions of optimum light and nutrient concentrations. Because of its low density in Lake Whatcom, none of the other statistical tools used *Ceratium* to identify late summer phytoplankton blooms. RIFFLE's ability to use both common and rare taxa is particularly useful for finding potential indicator species.

Principal components analysis was able to identify the major phytoplankton blooms; however, the results could easily be misinterpreted if importance was assigned to the individual species comprising each principal component rather than the trend that those species represent. For example, the winter diatom bloom was represented by *Melosira*, *Fragilaria* and *Tabellaria flocculosa* (Roth) Kutz. in the combined data set, but only by *Fragilaria* and *Melosira* in Basin 1 (see Table 1). This does not mean that *Tabellaria* was absent or rare in Basins 2 and 3; only that it accounted for less variation in the data sets for those basins. The interpretation of the summer phytoplankton blooms is even more difficult: the representative species are split into two groups in Basin 1, but only one group in the

TABLE 1

Principal components for Lake Whatcom phytoplankton, Basin 1 and all basins combined

Basin 1	Total	Species	Loading
PC-1	23%	*Dictyosphaerium* sp.	0.94
		Staurastrum sp.	0.93
		Aphanocapsa sp.	0.93
PC-2	16%	*Rhabdoderma* sp.	0.89
		Chroococcus sp.	0.87
		Oscillatoria sp.	0.85
PC-3	11%	*Fragilaria crotonensis*	0.93
		Melosira sp.	0.93
All Basins	**Total**	**Species**	**Loading**
PC-1	15%	*Dinobryon* sp.	0.790
		Coelsphaerium naegelianum	0.769
		Eudorina elegans Ehrenberg	0.774
		Unknown Greens	0.667
		Aphanocapsa sp.	0.542
PC-2	10%	*Melosira* sp.	0.905
		Fragilaria crotonensis	0.854
		Tabellaria flocculosa	0.847

combined data, and there is little overlap between the species in the different groups. While in some cases these results might lead to the discovery of an unknown pattern in the data, close inspection of the Lake Whatcom data does not support any such conclusion.

Correspondence analysis was more revealing. As can be seen from Fig. 8, there is a tendency for the COA score gradually to lessen during stratification, and swing rapidly back to its highest values immediately following turnover. This indicates that the large-scale gradient from a mixed to a stratified lake can be detected by correspondence analysis, and that the Lake Whatcom sample points successfully ordinated according to this trend. Basin 3, however, reveals that the presence of outliers can have a disastrous effect on this ordination technique. Gauch et al. (1977) make the same observation.

Hierarchical clustering proved ineffective in handling the Lake Whatcom phytoplankton data, typically resulting in highly unbalanced trees, whether squared Euclidean distance or cosine distance was used. The tree development was disastrously affected by outliers. Modification can be made to hierarchical clustering that improve its use for chronological samples. These modifications iclude: (a) transformations of the data matrix (normalization etc.), (b) the explicit removal of outliers from the data set during clustering, and (c) the imposition of a constraint to force temporally adjacent sample

points into the same clusters [see Allen et al. (1977); Legendre et al. (1985)]. However, these constraints seem excessively severe to us, and conceptual clustering provides an excellent alternative.

CONCLUSIONS

We conclude that limnological data sets are amenable to clustering and gradient analysis, with the proviso that care must be taken in the tools used. Principal components analysis was of some use in confirming water quality trends, in that it achieved a reduction in the redundancy of the data set by combining correlated parameters (such as temperature and pH) into a single component. However, principal components did not aid in the identification of large-scale patterns in the data, such as stratification. Further, used on data sets with many parameters (such as species lists) principal components provided only a marginal reduction in the complexity of the raw data. We found correspondence analysis to be superior to principal components for detecting large-scale gradients in the phytoplankton data from Lake Whatcom. This is consistent with the findings from theoretical studies of ordination (Kenkel and Orloci, 1986).

We believe that the results of this study, in conjunction with similar studies at other sites, will lead to an improvement in conventional biogeochemical modelling of limnological systems. Typically these models are lumped-parameter conceptual models, involving two crucial tasks. First, the model must be built on a small number of significant components, e.g. phosphorous, chlorophyll, phytoplankton or zooplankton, and, second, the gross, qualitative behavior of the lake must be understood in terms of changes in the states of these components (Scavia and Robertson, 1979, pp. 1–83). Conceptual clustering by RIFFLE helps by providing objective leads in both of these tasks: It provides an estimate, for each parameter, of how strongly the entire system is associated with that parameter; these estimates can guide the selection of components. It also provides a clustering of the samples of the lake system into states that may be significant parts of the evolution of the model.

Conceptual clustering was found to be consistently superior to hierarchical clustering. In clustering the physical chemical data, the presence of epilimnion and hypolimnion was clearly confirmed by our conceptual clustering algorithm. Hierarchical clustering did not isolate these clusters. In the phytoplankton set, a division into mixed and stratified communities was accomplished only by the conceptual clustering algorithm. This, together with the facts that (a) conceptual clustering makes fewer assumptions about the data than hierarchical clustering, and (b) it can handle incomplete and

mixed data sets without further assumptions or data subsetting, makes it a consistently superior tool for clustering.

ACKNOWLEDGEMENTS

We would like to thank the Institute for Watershed Studies, Western Washington University, and the City of Bellingham for their assistance with this project.

APPENDIX A

RIFFLE clustering

Clustering by the RIFFLE program (Matthews and Hearne, 1991) is a technique especially adapted to clustering ecological data. It is a partitional clustering algorithm: the data points are partitioned into clusters in a variety of ways, and the best such partition is selected as an appropriate clustering for the data. The 'best' clustering is one which maximizes the value of a fitness-measure (which evaluates the 'fitness' of the clusters to the data). The fitness-function used in RIFFLE estimates the accuracy of predictions in an imagined experiment, an experiment that uses the proposed cluster-membership of a sample to 'predict' whether that sample will have large or small values on its measured parameters. If a large number of these 'predictions' agree with the actual sample values, then the clustering fits. We use a nonparametric measure of fitness in the sense that predictions of numeric parameters are limited to the coarseness of the clustering. In a clustering into two groups, for example, only two values are predicted: 'high' values and 'low' values.

The quantitative measure of prediction accuracy used in RIFFLE is the proportional reduction in error, or *Guttman's* λ (Goodman and Kruskal, 1954). Suppose we wish to measure the fitness of a clustering into two groups, and we want to measure the accuracy of prediction for, say, a taxon t. Let a data point be represented by the vector x, with the point's value on parameter t be x_t. Let the two clusters be denoted by k_1 and k_2, and, for taxon t, let i_1 denote a 'high' value, and t_2 denote a 'low' value. (The best split value between 'high' and 'low' is also determined by the RIFFLE algorithm, but for concreteness we can assume the median is used.) A two-dimensional cross-tabulated frequency table, F, of the joint probabilities, is then built, where

$$F_{i,j} = \left| \{ x : x \in k_i \quad \text{and} \quad x_t = t_j \} \right|$$

i.e., F_{ij} is the number of times a sample is found which is in the ith cluster and has the jth value (high or low) of the taxa.

Under the usual statistical assumption that the distribution of sample points in F is representative of the distribution in the population, we can use F, and a knowledge of a sample's cluster, to predict the taxa count for that sample. If our sample is in cluster k_2, for example, our guess will be 'high' or 'low' depending on whether F_{21} or F_{22} has the larger value, and similarly if our sample is in cluster k_1.

If we do this for many samples, our total fraction of correct guesses C can be estimated to be:

$$C = \frac{\sum_i \text{Max}_j F_{ij}}{N}$$

where N is the total number of samples. The fraction on which we will be in error, then, will be $1 - C$. On the other hand, without a knowledge of a sample's cluster (and without using F), we can do no better in predicting 'high' or 'low' than 50% correct, on average (assuming a median split value). Our proportional reduction in error, therefore, using this clustering and its cross-classification table F, will be estimated to be:

$$\frac{\text{(Random Error)} - \text{(Clustered Error)}}{\text{Random Error}} = \frac{1/2 - (1 - C)}{1/2} = 2C - 1$$

The RIFFLE program searches over a large number of partitions of the data in order to maximize this proportional reduction in error for a large number of measured parameters. In other words, it searches for the one clustering (out of many) which is most closely associated with the measured parameters.

This algorithm has been implemented in Pascal and has been tested on a wide variety of computers and data sets (Matthews and Hearne, 1991).

APPENDIX B

Lake Whatcom water chemistry parameters sampled

Temperature	pH
Conductivity	Dissolved oxygen
Turbidity	Alkalinity
Secchi disk	Ammonia
Nitrate/Nitrite	Total nitrogen
Soluble reactive phosphate	Total phosphorus
Total organic carbon	Dissolved inorganic carbon
Chlorophyll *a*	

APPENDIX C

Lake Whatcom phytoplankton taxa list

Phylum: Chrysophyta	
Anomoeoneis serians (Breb. ex Kutz)	*Asterionella formosa* Hass.
Cyclotella compta (Ehr.) Kutz.	*Dinobryon* sp.
Fragilaria crotonensis Kitt.	*Melosira ambigua* (Grun.) o. Mull.
Melosira distans (Ehr.) Bethge.	*Navicula* sp.
Stephanodiscus sp.	*Synedra chaseana* (Thomas) Boyer
Synura sp.	*Tabellaria flocculosa* (Roth) Kutz
Phylum: Cyanophyta	
Anabaena sp.	*Anacystis* sp.
Aphanocapsa sp.	*Chroococcus* sp.
Coelosphaerium naegelianum Unger	*Gomphosphaeria lacustris* Chodat
Merismopedia tenuissima Lemmerman	*Microcystis aeruginosa* Kuetz.
Nostoc commune Vauch.	*Oscillatoria* sp.
Rhabdoderma sp.	*Schizothrix calcicola* (Ag.) Gom.
Phylum: Chlorophyta	
Dictyosphaerium sp.	*Eudorina elegans* Ehrenberg
Pandorina sp.	*Pediastrum duplex* Meyern.
Scenedesmus quadricauda (Turp.)	*Spondylosium* sp.
Staurastrum sp.	
Phylum: Pyrrhophyta	
Ceratium hirudinella (O.F. Muell.)	

REFERENCES

Allen, T.F.H., Bartell, S.M. and Koonce, J.F., 1977. Multiple stable configurations in ordination of phytoplankton community change rates. Ecology, 58: 1076–1084.

APHA, 1985. Standard Methods for the Examination of Water and Wastewater (16th Edition). American Public Health Association, American Water Works Association, and Water Pollution Control Federation, Washington, DC, 1268 pp.

Bayne, C.K., Beauchamp, J.J., Begovich, C.L. and Kane, V.E., 1980. Monte Carlo comparisons of selected clustering procedures. Pattern Recognition, 12: 51–62.

Ehinger, W.J., 1988. Plankton composition and temporal variation among the three basins of Lake Whatcom, Washington. Master's thesis, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA, 133 pp.

EPA, 1983. Methods of Chemical Analysis of Water and Wastes. EPA-600/4-79-020.

Fisher, D. and Langley, P., 1986. Conceptual clustering and its relation to numerical taxonomy. In: W.A. Gale (Editor), Artificial Intelligence and Statistics. Addison-Wesley, Reading, MA, pp. 77–116.

Gauch, H.G., Whittaker, R.H. and Wentworth, T.R., 1977. A comparative study of reciprocal averaging and other ordination techniques. J. Ecol., 65: 157–174.

Goodman, L.A. and Kruskal, W.H., 1954. Measures of association for cross classifications. J. Am. Stat. Assoc., 49: 732–745.

Kenkel, N.C. and Orloci, L., 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology, 67: 919–928.

Legendre, P., Dallot, S. and Legendre, L., 1985. Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. Am. Nat., 125: 257–288.

Lighthart, B., Kraft, G.F. and Flora, C.J., 1972. The limnology of Lake Whatcom. Tech. Rep. 15, Institute for Freshwater Studies, Western Washington State College, Bellingham, WA.

Lind, O.T., 1985. Handbook of Common Methods in Limnology. Kendall/Hunt, Dubuque, IA, 197 pp.

Lorenzen, C.J., 1967. Determination of chlorophyll and pheo-pigments: spectophotometric equations. Limnol. Oceanogr., 12: 343–346.

Ludwig, J.A. and Reynolds, J.F., 1988. Statistical Ecology. Wiley-Interscience, New York, 337 pp.

Matthews, G. and Hearne, J., 1991. Clustering without a metric. IEEE Trans. Pattern Anal. Machine Intell., 13 (forthcoming).

Pielou, E.C., 1984. The Interpretation of Ecological Data. A Primer on Classification and Ordination. Wiley-Interscience, New York, 263 pp.

Scavia, D. and Robertson, A., 1979. Perspectives on Lake Ecosystem Modeling. Ann Arbor Science, Ann Arbor, MI, 326 pp.

Dimensionality Reduction of Multivariate Ecotoxicological Data

Geoffrey Matthews

September 30, 1992

Contents

List of Figures

1 Introduction

The importance of dimensionality reduction in multivariate data analyses cannot be overstated. The visualization ability of humans is limited to two or three dimensions, while real-world ecological studies characteristically involve dozens to hundreds of dimensions. Scientific visualization today usually takes the form of allowing the scientist to interactively view 2D or 3D "slices," *i.e.* linear or nonlinear projections of his or her data. However, the large number of dimensions in most data makes the number of possible projections astronomically large. This combinatorical explosion prohibits random searching for patterns in the data by hand and sets limits on the utility of purely interactive visualization tools.

Dimensionality reduction is the attempt to give systematicity to the search for low-dimensional patterns in highly-dimensioned data. Many successful techniques already employed in ecology and toxicology, such as principal components, factor analysis, and multidimensional scaling, are simply tools for dimensionality reduction, tools which will tell the scientist *which* slices might be interesting to look at. These tools, however, were developed using simple, direct mathematical techniques, such as eigenvalues, which can be computed directly. Our work in artificial intelligence and nonparametric multivariate analysis has been shown to be a useful alternative in many ecological and toxicological analyses (Matthews et al., 1987; Matthews, 1988; Matthews et al., 1990b; Matthews and Matthews, 1990; Matthews et al., 1990a; Matthews and Hearne, 1991; Matthews et al., 1991a; Matthews et al., 1991b).

Artificial intelligence involving heuristic search promises to revise the process of dimensionality reduction. Heuristic search is a technique for efficiently searching a large space of possible solutions for the "best" one, one that maximizes one or more desirable properties. We will apply two kinds of heuristic search for achieving dimensionality reduction: clustering-directed hill-climbing and encoding neural nets.

Clustering-directed hill-climbing will seek a linear projection of the data space that maximizes clustering tendency. Maximizing the clustering tendency in a two-dimensional projection should give the user a good feel for some of the patterns in the fully-dimensioned data set. Since the result of the clustering-directed search process is a linear projection matrix, the interpretability of the projection vectors should be reasonably straightforward.

All linear projections, however, have the potential of losing information when the highly-dimensioned data is reduced to two dimensions. Encoding neural nets, however, do not attempt to find a linear two-dimensional projection matrix, but, rather, an encoding in two dimensions of all the information present in the fully-dimensioned data set. If successful, for example, such neural net encodings could reconstruct the entire data set, in ten or twenty dimensions, from only a two-dimensional representation of it. In a sense, neural net encodings seek to remove the redundancy from a data set, and reduce it to its essence. Plotting the data set under such a reduction is likely to reveal more patterns than in any linear projection.

2 Clustering-directed hill-climbing

Traditionally dimensionality reduction has taken the form of projections based on some form of the variance-covariance matrix of the data. In it's simplest form (principal components) the eigenvectors of the variance-covariance matrix (or the correlation matrix) are used for projections which maximize the variance. Other techniques are based on various normalizations and transformations of either the data matrix itself, or the square of the data matrix, or both. All of these techniques are limited by the assumption that the only computationally feasible projections are based on sums of squares, cross-products, and row and column marginal totals and extrema. The problems with this technique are illustrated in Figure 1.

The points in this artificial three-dimensional data set exhibit two cigar-shaped patterns. However, if this data is projected by a technique such as principal components, something like Figure 2 results, where the intuitive patterns are obscured. In contrast, if the data is projected as in Figure 3, then the patterns are obvious.

Our proposal is to use a heuristic search for a projection of maximum patterning. Patterning itself can be measured by many means; our preliminary work has shown that some measures of clustering tendency, in particular Guttman's lambda used with a quadrat histogram, are good indicators of intuitively obvious patterns in two-dimensional data (Guttman, 1941; Goodman and Kruskal, 1954; Jain and Dubes, 1988; Chen, 1992).

The basic data flow in the heuristic hill-climbing search algorithm is illustrated in Figure 4. First a random projection matrix is selected. Using

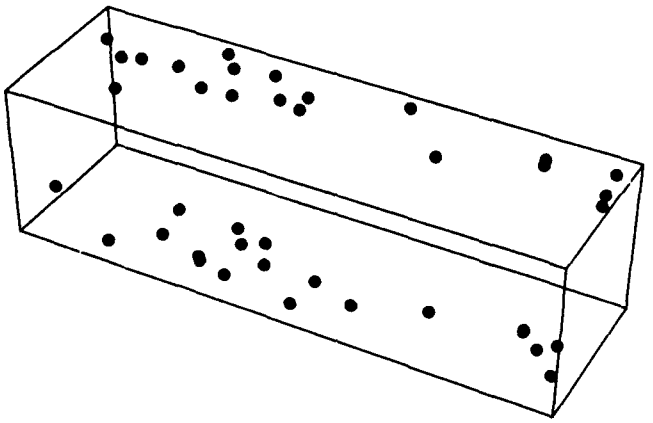


Figure 1: Hypothetical 3d data set.

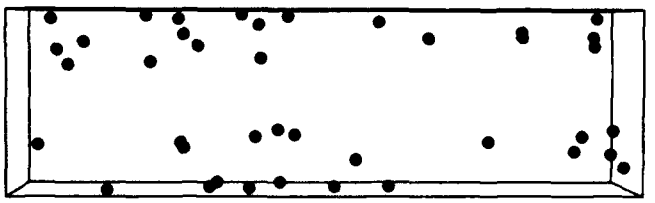


Figure 2: Hypothetical 3d data set projected on the components of maximum variance.

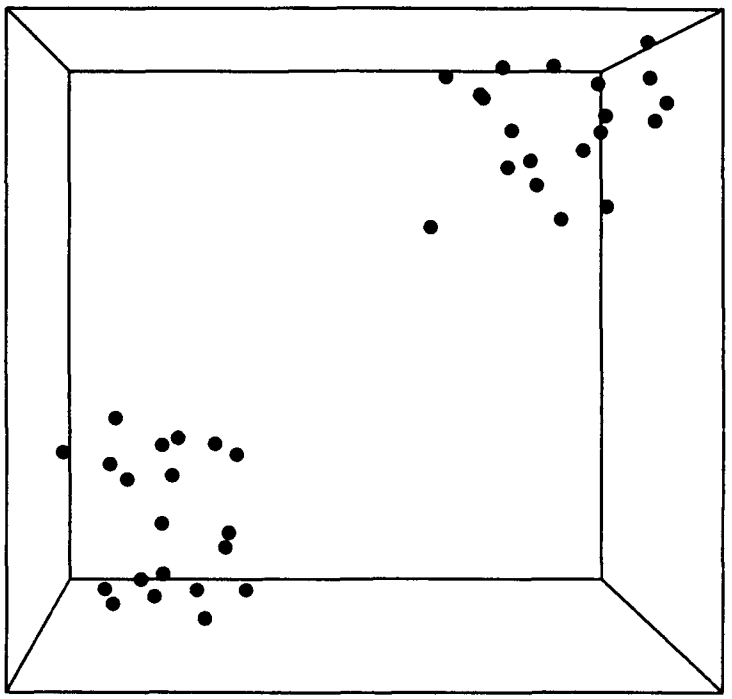


Figure 3: Hypothetical 3d data set projected to reveal patterning.

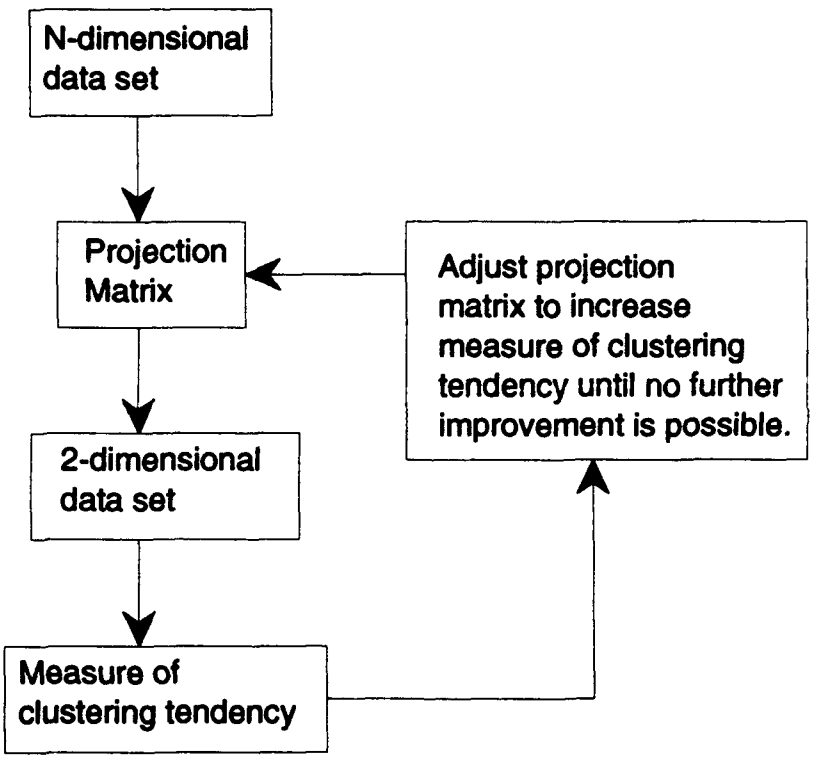


Figure 4: Basic steps in the hill-climbing search algorithm.

this matrix, the highly-dimensioned data are projected down to two dimensions. Clustering tendency is difficult to measure in highly-dimensioned data, but fairly easy and reliable in two dimensions (Jain and Dubes, 1988; Chen, 1992). This computed measure is then used to readjust the projection matrix: changes are made to the projection matrix over and over, as long as the computed two-dimensional measure of clustering tendency can be increased. When it reaches a maximum, the algorithm stops. In order to avoid the trap of local maxima, the whole process is repeated several times, starting with different (randomly selected) projection matrices each time.

3 Neural net encoding

An alternative view of dimensionality reduction rejects projections of the data matrix entirely. Instead, new parameters, which might have little to do with the original coordinates, are discovered. An analogy might be useful: If a tangled piece of string is laying on the ground, the "natural" coordinate for points on the string would be something based on "distance from one end" rather than a two-dimensional, x and y set of coordinates. A linear projection could not possibly discriminate all of the points, however. For example, any linear projection into one dimension of the points graphed in Figure 5 would put collapse some points to the same position. A principal components projection along the line of maximum variance, for instance, would map all points near the middle of the curve to the same point. This same phenomenon happens quite often for coenocline data, where the "arch effect" brings disparate points close together in the projection, and has been the subject of numerous attempts to overcome it (Gauch, 1982; Pielou, 1984).

Neural nets offer a proven alternative for discovering natural coordinates, as has already been demonstrated in (Saund, 1989). Data points can be encoded using a neural net that does not need to use linear projections from the n-dimensional space the data points come from. In Figure 6, the input and output layers are both n-dimensional, while the internal layer is only 2-dimensional. Using the data points, the neural net can be trained to encode the n-dimensional input into a 2-dimensional internal representation, which can be used by the decoding net to recover the n-dimensional data in the output layer. The input layer and output layer are compared to assure that not only is the input encoded into the internal layer, but the internal, two-

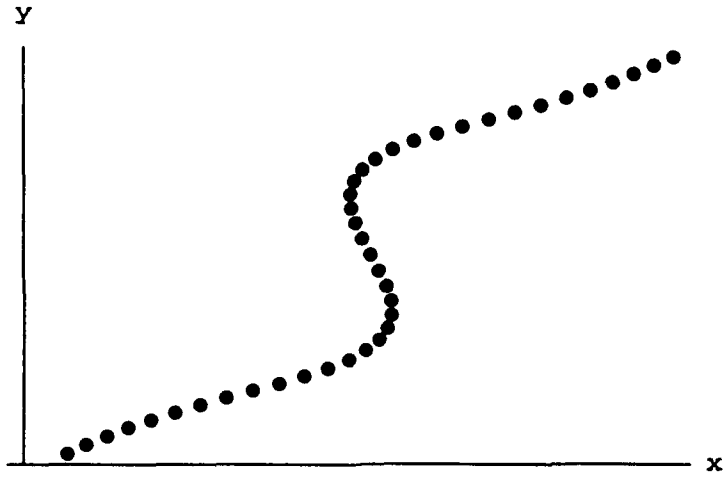


Figure 5: Two-dimensional data with a natural coordinate (distance from end) in one dimension. Linear projection would map points near the middle to the same point. A natural coordinate projection would put all the points along a single line.

dimensional layer by itself is sufficient to recover all of the input.

The figure is deliberately simplified. Each of the nodes in each layer in the figure is actually a scalar set of variables (Saund, 1989), and the input and output layers will usually have many more dimensions. Further, connections between layers actually go both ways, connecting "output" layers to internal and "input" layers, in order to implement backpropagation (Rumelhart et al., 1986).

References

Chen, C. (1992). The measurement of clustering tendency in machine learning. Master's thesis, Western Washington University, Bellingham Washington.

Gauch, Jr., H. G. (1982). *Multivariate Analysis in Community Ecology.* Cambridge University Press, Cambridge, MA.

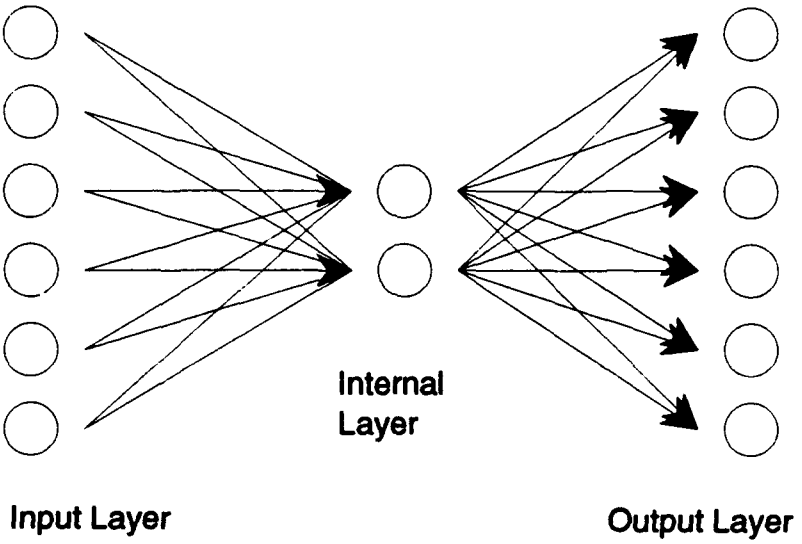


Figure 6: Simplified neural net architecture.

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49:732–764.

Guttman, L. (1941). An outline of the statistical theory of prediction. In Horst, P., editor, *The Prediction of Personal Adjustment*. S.S.R.C, New York.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ.

Matthews, G. and Hearne, J. (1991). Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175–184.

Matthews, G., Hearne, J., and Sugarman, P. (1987). Conceptual clustering in the analysis of environmental data sets. In *NOAA Conference on Artificial Intelligence Research in Environmental Science*, Boulder, Colorado.

Matthews, G. and Matthews, R. (1990). A model for describing community change. In *Pesticides in Natural Systems: How Can Their Effects Be Monitored? Proceedings of the Conference*, Environmental Research Laboratory/ORD, Corvallis Oregon. EPA 9109/9-91-011.

Matthews, G., Matthews, R., and Landis, W. (1990a). Applications of nonmetric clustering (nmc) to pattern analysis in aquatic toxicology. In *Society of Environmental Toxicology and Chemistry 11th Annual Meeting*, Arlington, Virginia.

Matthews, G. B. (1988). Clustering heterogeneous ecological data. In *Annual Conference, International Society for Ecological Modelling*, Davis, California.

Matthews, G. B., Matthews, R. A., and Hachmöller, B. (1991a). Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(11):2184–2190.

Matthews, R. A., Matthews, G. B., and Ehinger, W. J. (1991b). Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modelling*, 53:167–187.

Matthews, R. A., Matthews, G. B., and Hachmöller, B. (1990b). Ordination of benthic macroinvertebrates along a longitudinal stream gradient. In *Annual Conference, North American Benthological Society*, Blacksburg, Virginia.

Pielou, E. C. (1984). *The Interpretation of Ecological Data. A Primer on Classification and Ordination.* Wiley-Interscience, New York, NY.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol. 1*, pages 318–362. MIT Press, Cambridge, MA.

Saund, E. (1989). Dimensionality-reduction using connectionist networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(3):304–314.

Technical Report NCAA: Nonmetric Clustering and Association Analysis

Geoffrey Matthews*

November 3, 1992

Contents

1 Introduction

The fundamental assumption of nonmetric clustering and association analysis (NCAA) is the same as for other statistical tests: if the treatment had an effect, then data points taken from within one group will be more similar

*Computer Science Department, Western Washington University, Bellingham, Washington 98225

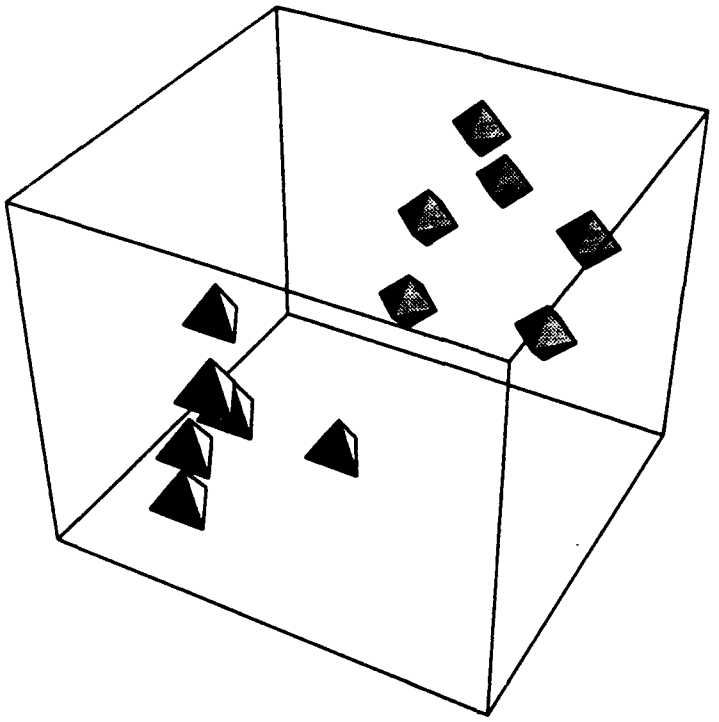


Figure 1: Clustered three-dimensional data points.

to each other than they will be to data points taken from a different group. Statistical tests differ primarily in how they measure similarity. The t-test, for instance, assumes that large differences in the mean values for the groups implies dissimilarity. The F-test, for another example, assumes that small variances within the groups implies similarity within them. Each of these attempts to determine whether the within-group-similarity is significantly larger than the between-group-similarity.

The strategy for multivariate statistical tests is intuitively illustrated in Figures 1 and 2. If we assume the tetrahedra represent the control group, and the octahedra represent the treatment group, and we plot the responses of three species to the treatment, we might get data that look like these pictures. It is intuitively obvious that in Figure 1, the within-group similarity is greater than the between-group similarity, while this fails to be the case in Figure 2. The problem, however, is in quantifying this intuition. With univariate

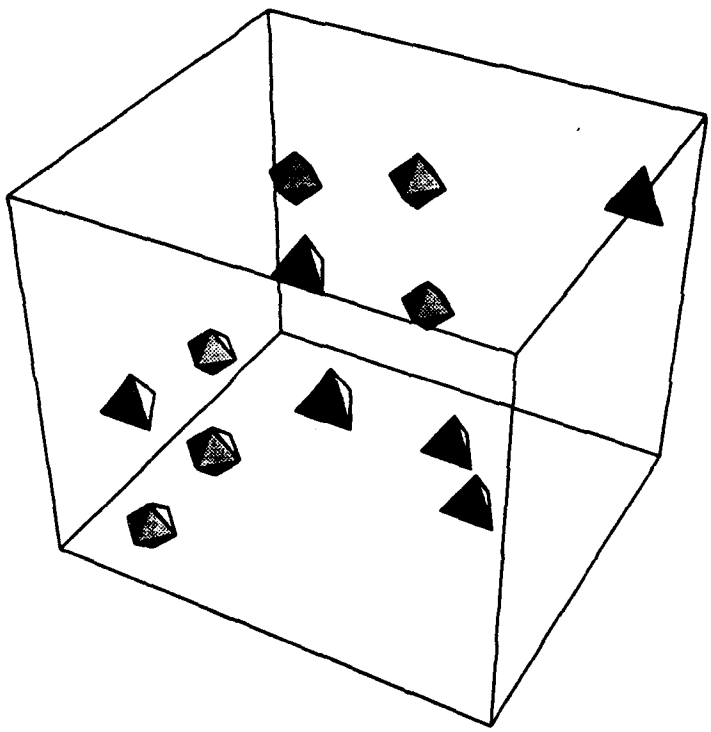


Figure 2: Unclustered three-dimensional data points.

tests, similarity means simply close numerical values. With multivariate data, however, there is no satisfyingly direct meaning for the concept of similarity.

In Section 2, we discuss definitions of multidimensional metrics. Section 3 introduces nonmetric clustering which relies on a more intuitively satisfying notion of similarity. In Section 4 we discuss the use of nonmetric clustering in a significance test to quantify the intuitive difference between a situation like Figure 1 and Figure 2.

2 N-dimensional distance metrics

The concept of similarity employed in statistical tests for multivariate data is crucial in determining the nature and suitability of the test. Typically, multivariate tests rely on a measure of similarity that involves a measure of vector separation between points. For example, the two points shown in Figure 3, have the following coordinates:

	x	y	z
Tetrahedron	2	3	5
Octahedron	8	9	1

The three numbers associated with each point can be viewed as three species, or as total algae, daphnia, and total nitrogen, *etc.* We can measure the similarity of one point to the other by a number of means. The Euclidean distance between the points, for example, would be:

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} = \sqrt{36 + 36 + 16} \approx 9.4$$

Or, we could draw two lines from the origin through the points, and then measure the size of the angle made by these lines. The cosine of this angle is also easily computed:

$$\frac{x_1x_2 + y_1y_2 + z_1z_2}{\sqrt{(x_1^2 + y_1^2 + z_1^2)(x_2^2 + y_2^2 + z_2^2)}} = \frac{16 + 27 + 5}{\sqrt{(4 + 9 + 25)(64 + 81 + 1)}} \approx 0.64$$

Since the cosine of an angle decreases monotonically as the angle increases, this number would be a measure of the similarity of the points, rather than the distance. These measures, and dozens more that can be found in the literature, essentially reduce two multidimensional points to a single number, giving a measure of their similarity, which I shall call a *similarity metric.*

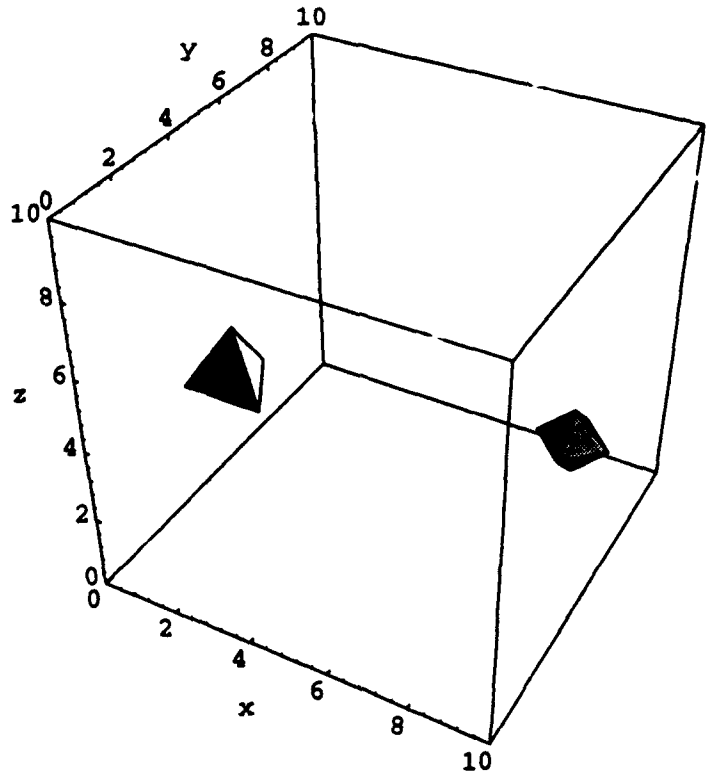


Figure 3: Two points in three-dimensional space.

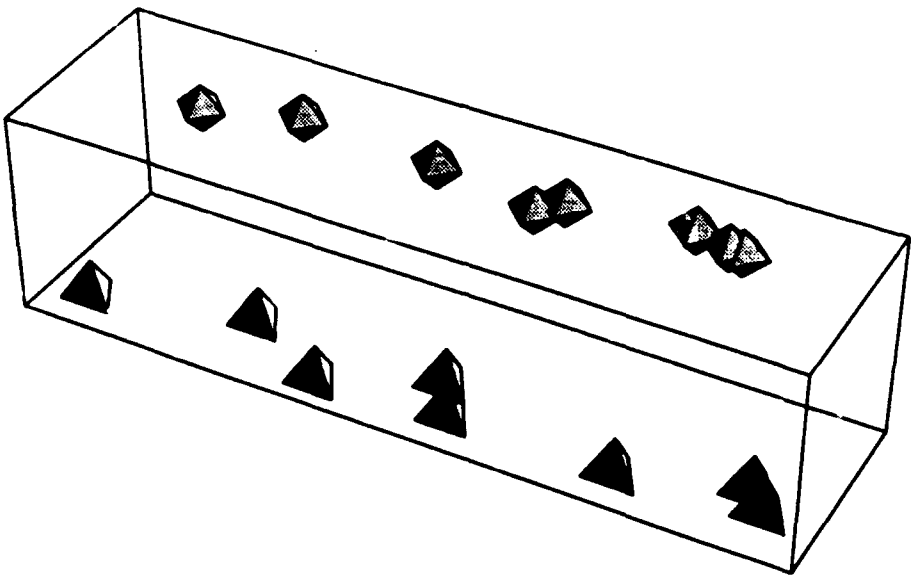


Figure 4: Two cigar-shaped clusters of points in three dimensions.

Similarity metrics are the basis for a number of multidimensional statistical tests. The simplest, by analogy to ANOVA, is to take the ratio of the average within-group similarity to the average between-group similarity (Smith et al., 1990). Clustering algorithms, which cluster points together into groups of similar points, usually judge similarity based on a distance metric (Jain and Dubes, 1988).

Similarity metrics are simple to compute, and are intuitively satisfying, but they have a number of problems associated with them. One of them is illustrated in Figure 4. Here we have two linear trends in the data, one exemplified by the tetrahedra running along the lower, front edge of the box, and one by the octahedra running along the upper, back edge of the box. A similarity metric, however, would be incapable of recognizing these patterns: two points at opposite ends of the same cigar are actually farther apart (and thus less similar) than two points at the same end of different cigars.

3 Nonmetric Clustering

In (Matthews and Hearne, 1991), we advanced a new notion of similarity in multivariate space which does not rely on a similarity metric. The essential feature of this nonmetric clustering is that the quality of the clustering is not based on a point-by-point, within *vs.* between, comparison of similarity. Instead, we take it as obvious that the "ideal" clusters would be strongly associated with the responses of a number of species. If most of the species are numerous in one group of points, but low in another group, then those groups make natural clusters.

The strategy for assigning points to clusters is illustrated in Figure 5, where we have two clusters of points, labelled "A" and "B". The question is, into which cluster do we put the point labelled "U"? The U-point should be assigned to cluster A, but *not* because it is closer to the A-points. Indeed, in the figure, it is difficult to judge whether U is closer, on average, to the A's or to the B's. But in both the x-axis projection and the y-axis projection, U is clearly in the midst of the A's, not the B's. If the U-point is assigned to the A-cluster, then the projections look like this:

AAAAAABBBBB

But if the U-point is assigned to the B-cluster, then both projections look like this:

AAABAABBBBB

A measure of association between the cluster labels (the A's and B's) and the x and y axes will show better association with U in the A cluster than with U in the B cluster.

The "ideal" clustering will show a strong association like this with *all* of the data parameters, *i.e.* with every species in the test. This ideal will usually be unachievable, but still we can rank clusterings as better or worse in accord with the strength of their association with most of the species. Looking back at Figure 4, for instance, we can see that the illustrated clustering, into tetrahedra and octahedra, has a perfect association in two dimensions (the two short dimensions), but poor in the third (the long dimension). A nonmetric clustering of this data will not break the cigars.

In addition to ranking clusterings, and thus being able to look for a "best" clustering, this approach will also rank parameters with respect to the final, best clustering. The more strongly a parameter is associated with the final clustering, the more "important" that parameter is to the final clustering.

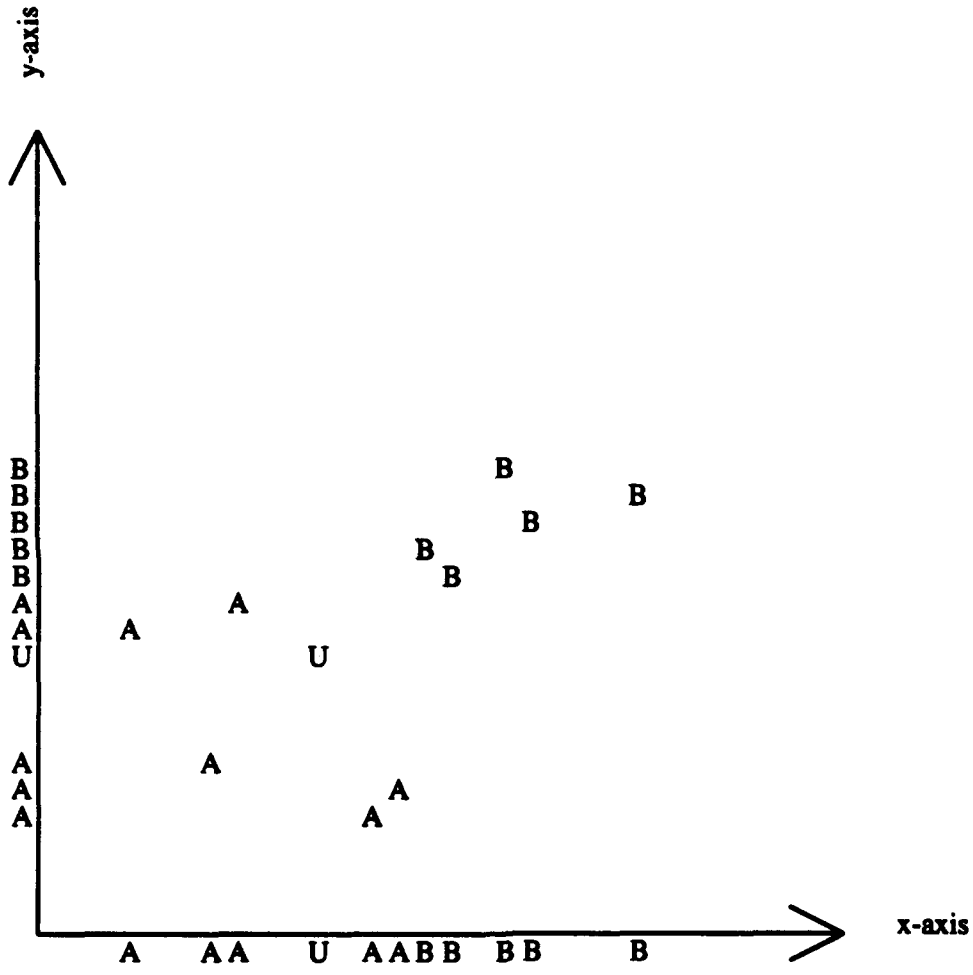


Figure 5: Clustered points in two dimensions with their one-dimensional projections. Point U goes in cluster A because both the x-axis projection and the y-axis projection would show improved association with the clusters.

The computer program RIFFLE (Matthews and Hearne, 1991) was built to find a clustering of data points that showed the strongest association possible with the largest number of dimensions, and then report both the clustering and the relative importance of each dimension to that clustering. This clustering methodology has a number of advantages over conventional methods:

- It does not combine counts from dissimilar taxa by means of sums of squares, or other *ad hoc* mathematical techniques.
- It does not require transformations of the data, such as normalizing the variance.
- It works without modification on incomplete data sets.
- It can work without further assumptions on different data types (*e.g.*, species counts or presence/absence data).
- Significance of a taxon to the analysis is not dependent on the absolute size of its count, so that taxa having a small total variance, such as rare taxa, can compete in importance with common taxa, and taxa with a large, random variance will not automatically be selected, to the exclusion of others.
- It provides an integral measure of "how good" the clustering is, *i.e.* whether the data set differs from a random collection of points.
- It can, in some cases, identify a subset of the taxa that serve as reliable indicators of the physical environment. In our research the indicator species selected by RIFFLE often proved to be more reliable than indicators based on a linear discriminant (Matthews et al., 1991a; Matthews et al., 1991b).

The major disadvantage of the RIFFLE program is that, in order to find a clustering of the data points with the desirable qualities listed above, a massive search through thousands of potential clustering candidates is made before settling on the "right" one. Even after this search, there is no guarantee that RIFFLE finds the optimal clustering, in the sense outlined above. However, in our research, RIFFLE does find an excellent clustering in a reasonable amount of time.

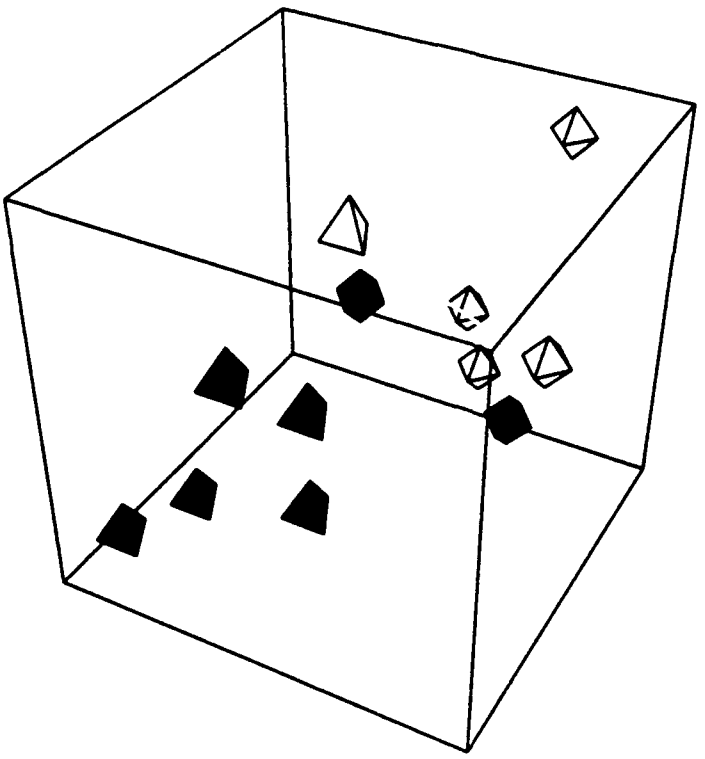


Figure 6: Multivariate points from two treatment groups, marked by octahedra and tetrahedra, and clustered into two groups, marked by light and dark coloring.

4 Association Analysis

Clustering the data points from a multivariate test is only half of the game, however. A significance test of the effect of the treatment is still needed. Our approach to this is illustrated in Figure 6. Each point has both a treatment group (marked by the shape of the polyhedron) and a cluster (marked by the coloring of the polyhedron). Bear in mind that the points were assigned to clusters independently of which treatment group the point came from. It is only *after* the clustering by RIFFLE is complete that the association between groups and clusters is considered. The clustering itself is completely blind to treatment groups.

Further, the use of *nonmetric* clustering is not essential to this stage of the analysis. The association between clusters and treatment groups could be carried out after any clustering methodology, such as hierarchical or k-means clustering.

Now that each point has both a cluster and a group, the association between clusters and groups can be evaluated in a contingency table format. For instance, the points in Figure 6 would fill out the following table:

	Tetrahedra	Octahedra
Light	1	4
Dark	5	2

Under the null hypothesis that the treatment group has no effect on the data, the points in one treatment group would be just as likely to be in one cluster as another, and a uniform distribution of points in the contingency table would be expected. The Pearson χ^2 for the table can then be computed (Fienberg, 1985), to judge the significance of the effect (*i.e.* the probability, under the null hypothesis, of obtaining a χ^2 value at least as large as that of the observed table). Using this table, we get a value for χ^2 as follows:

$$\begin{aligned}\chi^2 &= \sum_{i,j} \frac{(N_{ij} - n_{ij})^2}{n_{ij}} \\ &= \frac{(1-2.5)^2}{2.5} + \frac{(4-2.5)^2}{2.5} + \frac{(5-3.5)^2}{3.5} + \frac{(2-3.5)^2}{3.5} \\ &\approx 3.09\end{aligned}$$

where N_{ij} is the actual cell count and n_{ij} is the expected cell count. With one degree of freedom (for a 2×2 table), this value can be looked up in a table of χ^2 probabilities to tell us that this (hypothetical) experiment shows a significant effect at the 90% level, but not the 95% level. Alternatively, a randomization or permutation test could be used to judge significance (Noreen, 1989).

Much toxicological testing uses four treatment groups, rather than two, but the strategy is the same. The data are clustered into four clusters, a (4×4) contingency table of treatment groups *vs.* clusters is assembled, and the significance of the effect is measured from the contingency table.

The group-cluster contingency table, however, can be used for more than simple hypothesis testing. The contents of the table can be examined to determine whether, for example, all four treatment groups were distinct, or

only one or two of them. More sensitive measures of association in contingency tables, such as Guttman's λ or entropy (Goodman and Kruskal, 1954; Goodman and Kruskal, 1959; Goodman and Kruskal, 1963; Goodman and Kruskal, 1972) could also be used, to judge not *whether* an effect occurred, but *how strong* the effect was.

5 Conclusion

Nonmetric clustering and association analysis (NCAA) is a tool for evaluating the effects of treatments on multivariate systems. Because it is based on nonmetric clustering, the tool can be used on "messy" data, with missing points or with variates that do not obey assumptions of normality and homoscedasticity. In addition to evaluating the strength of the effect, the tool also provides insight into which of the variates are most strongly associated with the effect. It is also a "blind" test, in that the clustering is done independently of the treatment groups. This is useful in screening experiments for unlooked-for effects, such as edge-effects in mesocosms.

References

Fienberg, S. E. (1985). *The Analysis of Cross-Classified Categorical Data.* MIT Press, Cambridge, MA.

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49:732–764.

Goodman, L. A. and Kruskal, W. H. (1959). Measures of association for cross classifications ii: further discussion and references. *Journal of the American Statistical Association*, 54:123–163.

Goodman, L. A. and Kruskal, W. H. (1963). Measures of association for cross classifications iii: approximate sampling theory. *Journal of the American Statistical Association*, 58:310–364.

Goodman, L. A. and Kruskal, W. H. (1972). Measures of association for cross classifications iv: simplification of asymptotic variances. *Journal of the American Statistical Association*, 67:415–421.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ.

Matthews, G. and Hearne, J. (1991). Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175–184.

Matthews, G. B., Matthews, R. A., and Hachmöller, B. (1991a). Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(11):2184–2190.

Matthews, R. A., Matthews, G. B., and Ehinger, W. J. (1991b). Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modelling*, 53:167–187.

Noreen, E. W. (1989). *Computer Intensive Methods for Testing Hypotheses*. Wiley-Interscience, New York, NY.

Smith, E. P., Pontasch, K. W., and Cairns, Jr., J. (1990). Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.*, 24(4):507–514.

Nonmetric Clustering and Projections from N-dimensional Ecosystem Space: New Approaches for the Analysis of Ecological Datasets

USEPA, Corvallis, Oregon

Geoffrey Matthews*

January 26, 1993

*Computer Science Department, Western Washington University, Bellingham, Washington 98225, matthews@cs.wwu.edu. This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

Contents

List of Figures

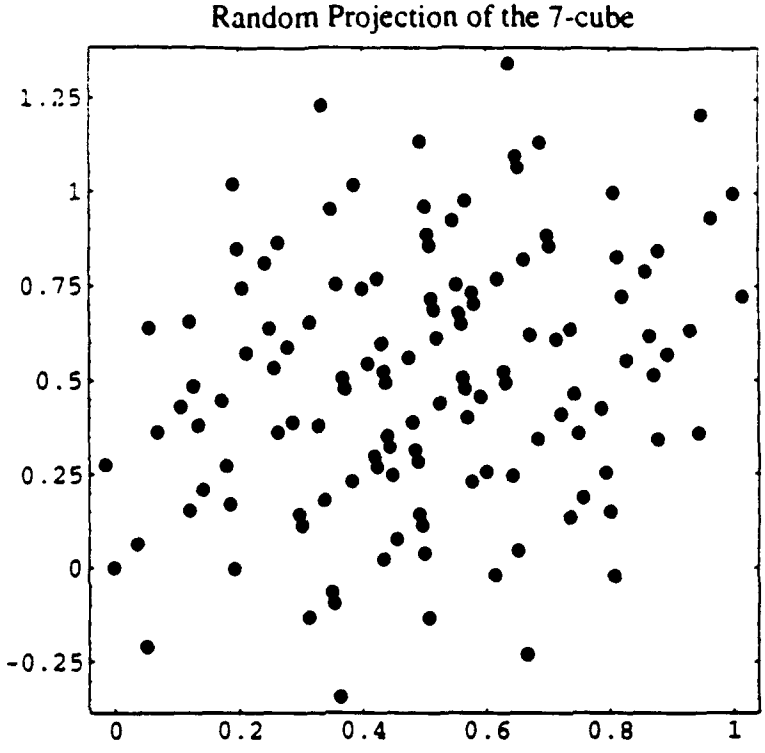


Figure 1: Corners of the 7-cube projected into two dimensions.

1 Introduction—The Curse of Dimensionality

High-dimensional space has a great deal of room. For example, if you have a large number of points distributed uniformly over a unit-radius sphere in 10 dimensions, then a sphere of radius 0.74 can be expected to contains only 5% of the points (Huber, 1985). For another example, consider points regularly distributed on the corners of a 7-dimensional cube, *i.e.* all points with coordinates which are either 0 or 1, for example; $(0,0,0,0,0,0,0)$, $(0,0,0,0,1,0,1)$, and $(1,1,1,1,0,0,0)$. Visualizing the structure of these 128 points can be difficult with projections. A random projection is likely to look like Figure 1, practically indistinguishable from a single-population, multivariate normal distribution, while a "good" projection will look more like Figure 2, and might suggest four normal subpopulations. Which pattern is real?

It does no good to suggest we look at "all" projections. Even projections onto pairs of the original axes would involve onerous work. For ten dimensions, there are 45 projections into two axes, and 120 projections into three axes. But these are just the projections into original coordinates: more general projections (such as the kinds provided by Principal Components Analysis or Correspondence Analysis) involve projecting at arbitrary angles to the original coordinates. Asimov has proposed a dynamic look at all angles, a display that slowly rotates through all possible projections, monitored by the user so that if any "interesting" patterns show up (such as the ones in Figure 2) the user can stop the display and investigate (Asimov, 1985). However, rotating at

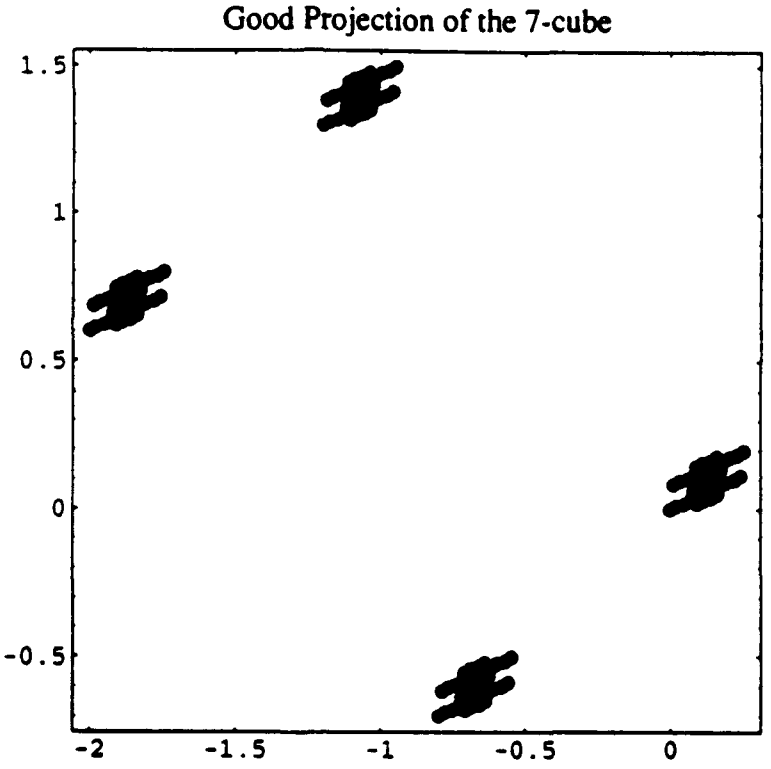


Figure 2: Corners of the 7-cube projected into two dimensions, in a different direction.

about 10° per second, a reasonable speed for careful observation, a grand tour of only four dimensions would take about three hours (Huber, 1985). In field studies, where the number of species and physical/chemical parameters recorded can easily exceed 20 or 30, and hence the dimensionality of the sample points will also be in excess of 20, an exhaustive study of all projections is clearly out of the question.

In addition to the sheer size of high-dimensional space, there is also the problem of "nuisance" variables—parameters which have a large variance but little information. Not only do they add to the general noisiness of the picture, but their mere presence adds to the dimensionality of every point, and thus exponentially complicates the analysis.

A number of approaches to this problem have been attracting increasing attention recently. We present here some of the background to techniques that have proved useful to us in both toxicology and ecology. First, we discuss in some detail nonmetric clustering and association analysis, a technique developed to overcome many of the problems associated with traditional clustering algorithms. Then, we present a brief survey of some other techniques and challenging problems that we see as becoming increasingly important in future research.

2 Nonmetric Clustering and Association Analysis

The fundamental assumption of nonmetric clustering and association analysis (NCAA) is the same as for other statistical tests: if the treatment had an effect, then data points taken from within one group will be more similar to each other than they will be to data points taken from a different group. Statistical tests differ primarily in how they measure similarity. The t-test, for instance, assumes that large differences in the mean values for the groups implies dissimilarity. The F-test, for another example, assumes that small variances within the groups implies similarity within them. Each of these attempts to determine whether the within-group-similarity is significantly larger than the between-group-similarity.

The strategy for multivariate statistical tests is intuitively illustrated in Figures 3 and 4. If we assume the tetrahedra represent the control group, and the octahedra represent the treatment group, and we plot the responses of three species to the treatment, we might get data that look like these pictures. It is intuitively obvious that in Figure 3, the within-group similarity is greater than the between-group similarity, while this fails to be the case in Figure 4. The problem, however, is in quantifying this intuition. With univariate tests, similarity means simply close numerical values. With multivariate data, however, there is no satisfyingly direct meaning for the concept of similarity.

In Section 2.1, we discuss definitions of multidimensional metrics. Section 2.2 introduces nonmetric clustering which relies on a more intuitively satisfying notion of similarity. In Section 2.3 we discuss the use of nonmetric clustering in a significance test to quantify the intuitive difference between a situation like Figure 3 and Figure 4.

2.1 *N*-dimensional distance metrics

The concept of similarity employed in statistical tests for multivariate data is crucial in determining the nature and suitability of the test. Typically, multivariate tests rely on a measure of similarity that involves a measure of vector separation between points. For example, the two points shown in Figure 5, have the following coordinates:

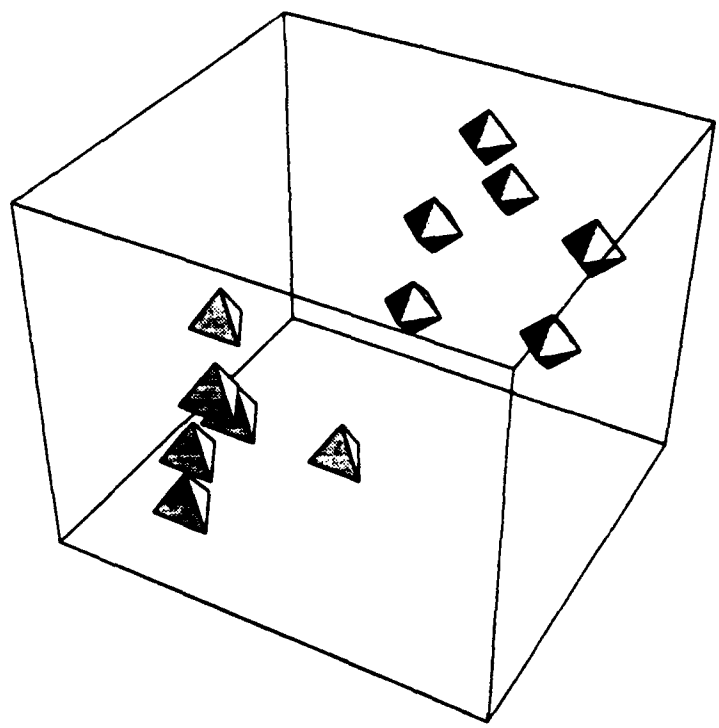


Figure 3: Clustered three-dimensional data points.

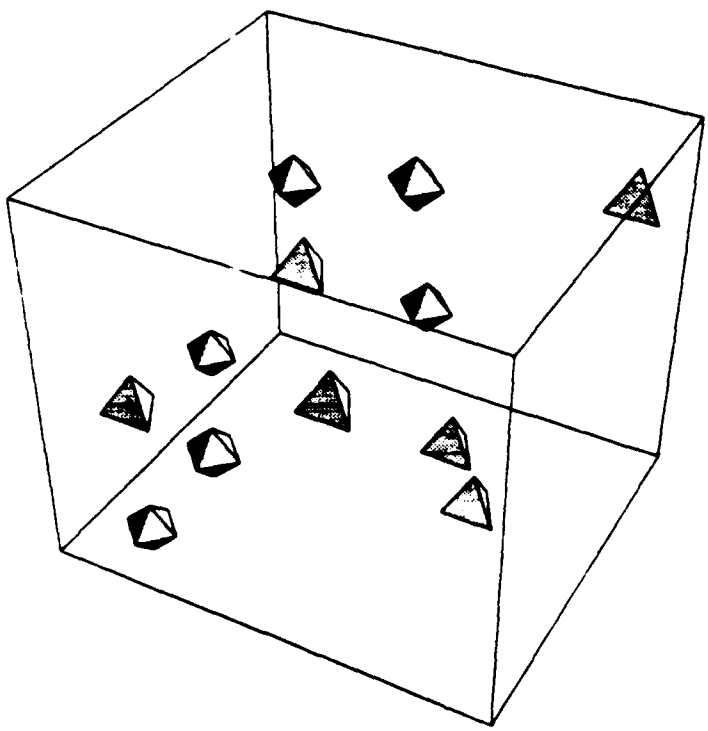


Figure 4: Unclustered three-dimensional data points.

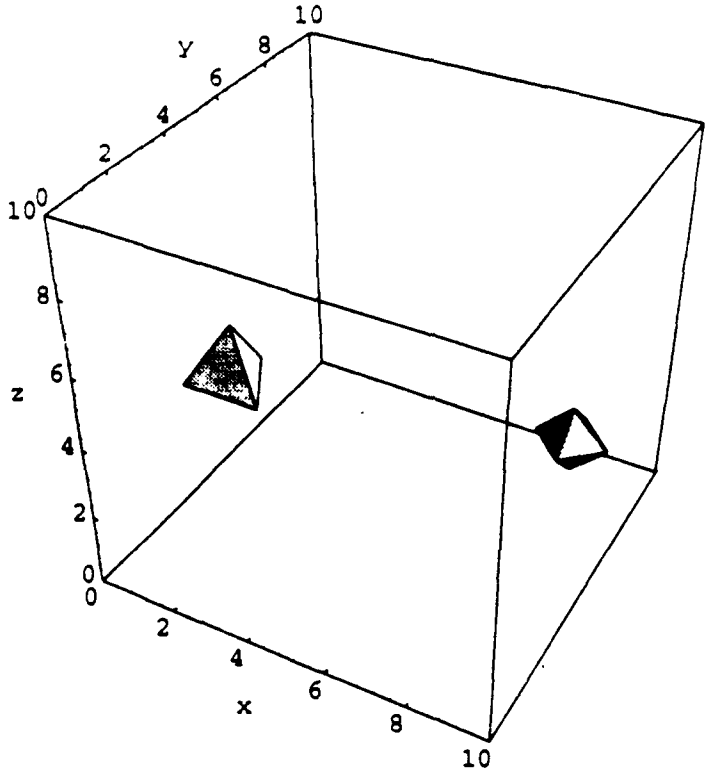


Figure 5: Two points in three-dimensional space.

	x	y	z
Tetrahedron	2	3	5
Octahedron	8	9	1

The three numbers associated with each point can be viewed as three species, or as total algae, daphnia, and total nitrogen, *etc.* We can measure the similarity of one point to the other by a number of means. The Euclidean distance between the points, for example, would be:

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} = \sqrt{36 + 36 + 16} \approx 9.4$$

Or, we could draw two lines from the origin through the points, and then measure the size of the angle made by these lines. The cosine of this angle is also easily computed:

$$\frac{x_1 x_2 + y_1 y_2 + z_1 z_2}{\sqrt{(x_1^2 + y_1^2 + z_1^2)(x_2^2 + y_2^2 + z_2^2)}} = \frac{16 + 27 + 5}{\sqrt{(4 + 9 + 25)(64 + 81 + 1)}} \approx 0.64$$

Since the cosine of an angle decreases monotonically as the angle increases, this number would be a measure of the similarity of the points, rather than the distance. These measures, and dozens more that can be found in the literature, essentially reduce two multidimensional points to a single number, giving a measure of their similarity, which I shall call a *similarity metric*. Similarity metrics are the basis for a number of multidimensional statistical tests. The simplest, by analogy to ANOVA, is to take the ratio

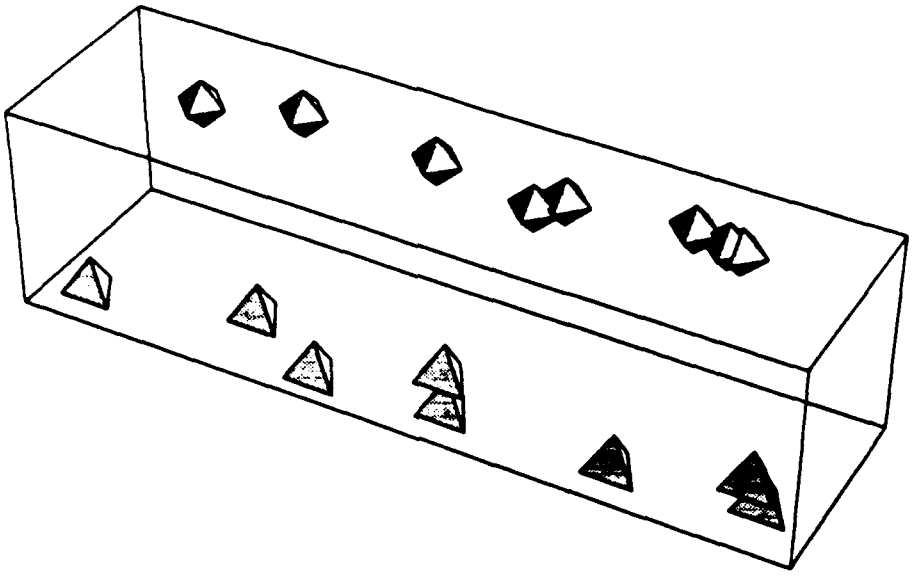


Figure 6: Two cigar-shaped clusters of points in three dimensions.

of the average within-group similarity to the average between-group similarity (Smith et al., 1990). Clustering algorithms, which cluster points together into groups of similar points, usually judge similarity based on a distance metric (Jain and Dubes, 1988).

Similarity metrics are simple to compute, and are intuitively satisfying, but they have a number of problems associated with them. One of them is illustrated in Figure 6. Here we have two linear trends in the data, one exemplified by the tetrahedra running along the lower, front edge of the box, and one by the octahedra running along the upper, back edge of the box. A similarity metric, however, would be incapable of recognizing these patterns: two points at opposite ends of the same cigar are actually farther apart (and thus less similar) than two points at the same end of different cigars.

2.2 Nonmetric Clustering

In (Matthews and Hearne, 1991), we advanced a new notion of similarity in multivariate space which does not rely on a similarity metric. The essential feature of this nonmetric clustering is that the quality of the clustering is not based on a point-by-point, within *vs.* between, comparison of similarity. Instead, we take it as obvious that the "ideal" clusters would be strongly associated with the responses of a number of species. If most of the species are numerous in one group of points, but low in another group, then those groups make natural clusters.

The strategy for assigning points to clusters is illustrated in Figure 7, where we have two clusters of points, labelled "A" and "B". The question is, into which cluster do we put the point labelled "U"? The U-point should be assigned to cluster A, but *not* because it is closer to the A-points. Indeed, in the figure, it is difficult to judge whether U is closer, on average, to the A's or to the B's. But in both the x-axis projection and

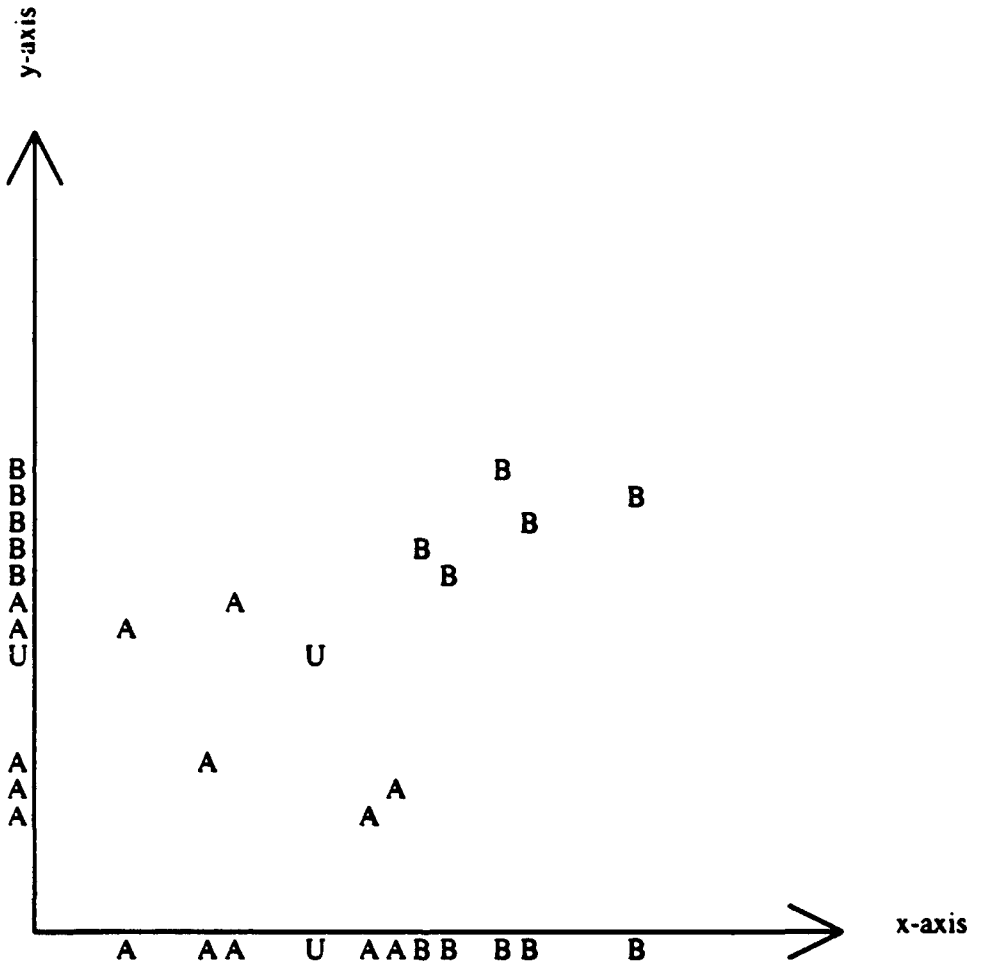


Figure 7: Clustered points in two dimensions with their one-dimensional projections. Point U goes in cluster A because both the x-axis projection and the y-axis projection would show improved association with the clusters.

the y-axis projection. U is clearly in the midst of the A's, not the B's. If the U-point is assigned to the A-cluster, then the projections look like this:

AAAAAABBBBB

But if the U-point is assigned to the B-cluster, then both projections look like this:

AAABAABBBBB

A measure of association between the cluster labels (the A's and B's) and the x and y axes will show better association with U in the A cluster than with U in the B cluster.

If all parameters were ordered, as numeric data is, a simple nonmetric test of association, such as the runs test, with each axis would suffice. However, many parameters are categorical and a more general test is called for. Nonmetric clustering measures the association between a clustering (which, itself, is a categorical variable) and another categorical variable by means of a cross-tab test. A frequency table of cluster-number *vs.* categorical-value is set up, and the number of data points in each cell is counted in order to measure the association between cluster and variable. The most famous cross-tab test is the χ^2 test, but the χ^2 test has some undesirable properties when it comes to interpretation, and instead nonmetric clustering, in its current form, uses Guttman's λ to measure the association in the table (Goodman and Kruskal, 1954; Goodman and Kruskal, 1959; Goodman and Kruskal, 1963; Goodman and Kruskal, 1972).

The frequency table approach works well for categorical variables, but what about numeric variables? Nonmetric clustering takes a pragmatic approach to these: *if* the data are going to be adequately described by the clustering (and the whole method is predicated on its possible success), say into three clusters, then there are really only three values of a numeric parameter to consider: low, middle, and high. All other variations in a numeric parameter will be assumed due to variance within the clusters. Accordingly, we can divide up the range of a numeric parameter into three parts. We can do this nonmetrically by simply choosing the 33.3 and 66.6 percentile points. This is illustrated in Figure 8, where we have clustered Fisher's famous "iris" data, and plotted the results in two of the four original dimensions. The gray lines indicate where the data divide up into three sections, along each axis. The points are clustered so as to achieve a maximum association with each axis, where the numeric value on each axis is simply converted to a categorical value based on these (nonuniform, nonmetric, data-driven) quadrats.

Of course, there are usually more than two dimensions involved. Even in the simple "iris" data there are four dimensions. These are plotted a scatterplot matrix in Figure 9, where, it can be seen, there are a lot of cells to consider. The clustering of the points (their shape and color, in the figure) is selected to maximize, so far as possible, the association between cluster and each axis. Intuitively, an attempt is made to make each cell in Figure 9 as homogeneous as possible. It might seem easy to make a single cell completely homogeneous, by simply coloring (clustering) all points within it the same. However, *each* data point shows up in *every* scatterplot in Figure 9, and recoloring one cell to make it better may in fact make other cells *less* homogeneous.

The "ideal" clustering will show a strong association with *all* of the data parameters, *e.g.* with every species in the test, and result in a uniform cluster within each cell of the "quantile quadrats". This ideal will usually be unachievable, but still we can rank clusterings as better or worse in accord with the strength of their association with most of the species. Looking back at Figure 6, for instance, we can see that the illustrated clustering, into tetrahedra and octahedra, has a perfect association in two dimensions (the two short dimensions), but poor in the third (the long dimension). A nonmetric

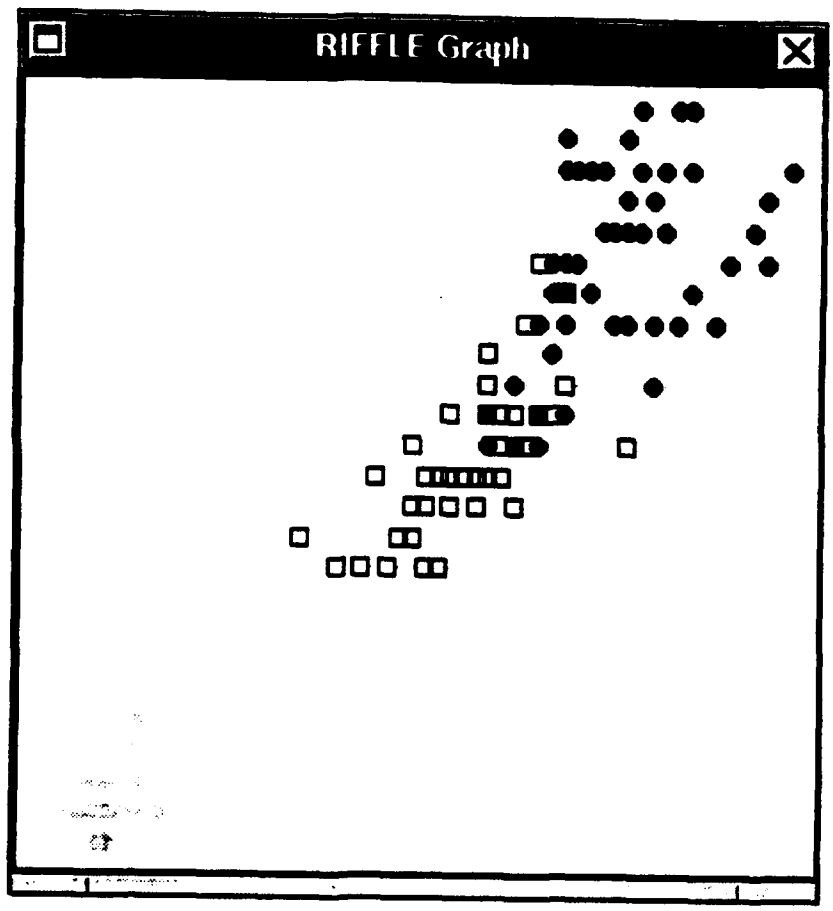


Figure 8: Nonmetric clustering of Fisher's "iris" data set. Nonmetric quadrats are shown with gray lines.

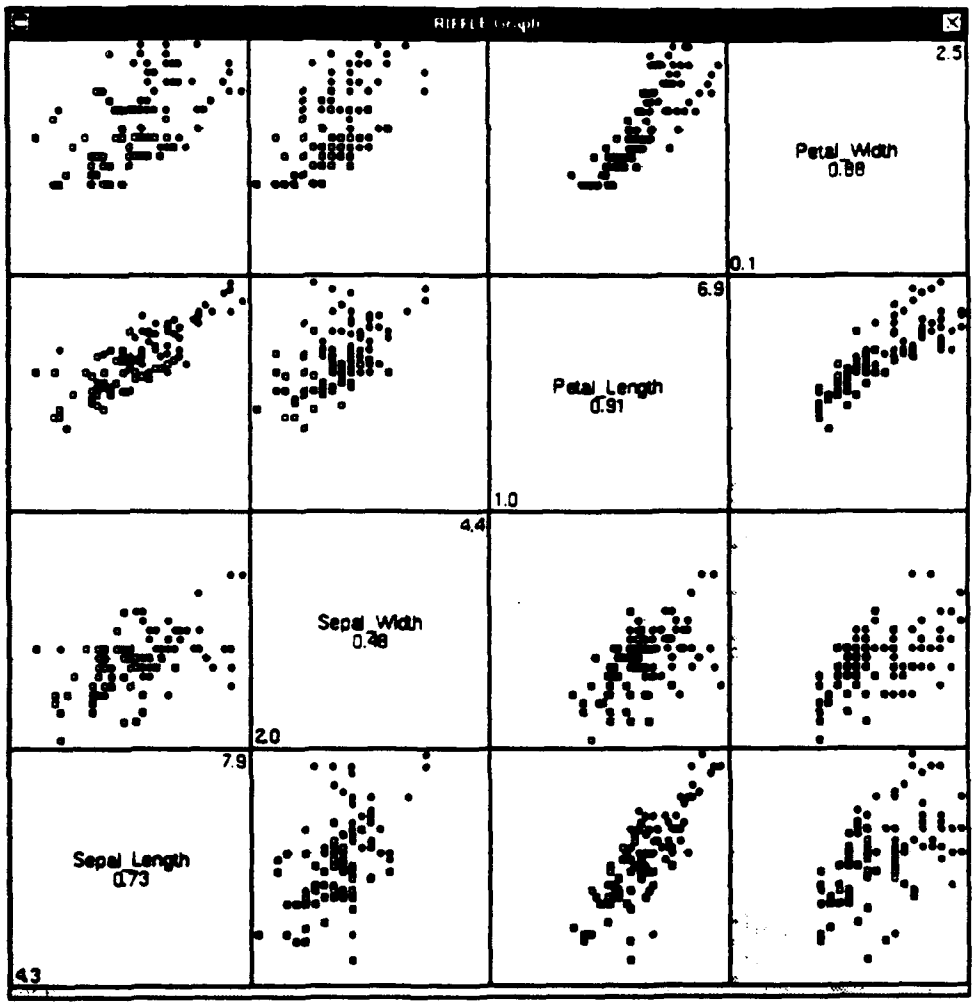


Figure 9: The same nonmetric clustering of Fisher's "iris" data set, shown in all four dimensions. Nonmetric quadrats are again shown with gray lines.

clustering of these data will not break the cigars, and will show a good association with two of the three dimensions.

Therefore, in addition to ranking clusterings, and thus being able to look for a "best" clustering, this approach will also rank parameters with respect to the final, best clustering. The more strongly a parameter is associated with the final clustering, the more "important" that parameter is to the final clustering.

The computer program RIFFLE (Matthews and Hearne, 1991) was built to find a clustering of data points that showed the strongest association possible with the largest number of dimensions, and then report both the clustering and the relative importance of each dimension to that clustering. This clustering methodology has a number of advantages over conventional methods:

- It does not combine counts from dissimilar taxa by means of sums of squares, or other *ad hoc* mathematical techniques.
- It does not require transformations of the data, such as normalizing the variance.
- It works without modification on incomplete data sets.
- It can work without further assumptions on different data types (*e.g.*, species counts or presence/absence data).
- Significance of a taxon to the analysis is not dependent on the absolute size of its count, so that taxa having a small total variance, such as rare taxa, can compete in importance with common taxa, and taxa with a large, random variance will not automatically be selected, to the exclusion of others.
- It provides an integral measure of "how good" the clustering is, *i.e.* whether the data set differs from a random collection of points.
- It can, in some cases, identify a subset of the taxa that serve as reliable indicators of the physical environment. In our research the indicator species selected by RIFFLE often proved to be more reliable than indicators based on a linear discriminant (Matthews et al., 1991a; Matthews et al., 1991b).

The major disadvantage of the RIFFLE program is that, in order to find a clustering of the data points with the desirable qualities listed above, a massive search through thousands of potential clustering candidates is made before settling on the "right" one. Even after this search, there is no guarantee that RIFFLE finds the optimal clustering, in the sense outlined above. However, in our research, RIFFLE does find an excellent clustering in a reasonable amount of time.

2.3 Association Analysis

Clustering the data points from a multivariate test is only half of the game, however. A significance test of the effect of the treatment is still needed. Our approach to this is illustrated in Figure 10. Each point has both a treatment group (marked by the shape of the polyhedron) and a cluster (marked by the coloring of the polyhedron). Bear in mind that the points were assigned to clusters independently of which treatment group the point came from. It is only *after* the clustering by RIFFLE is complete that the

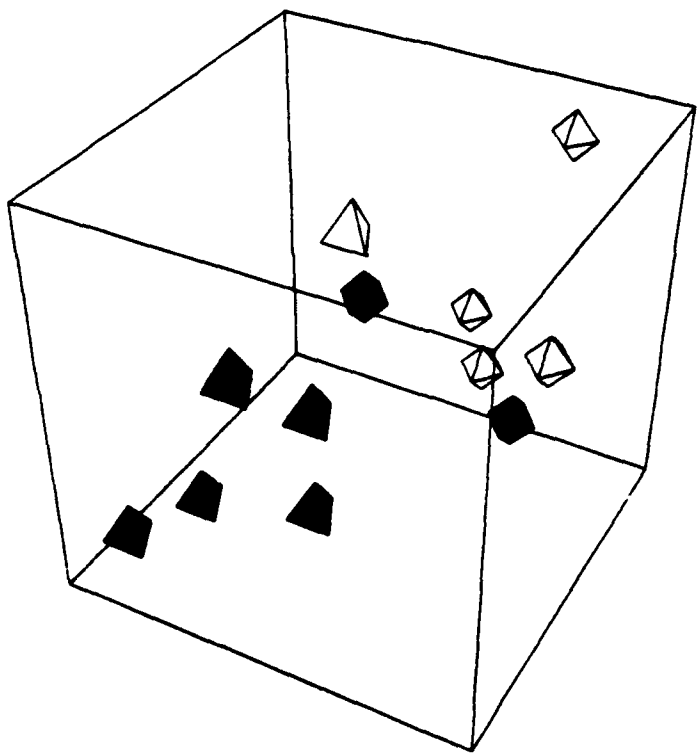


Figure 10: Multivariate points from two treatment groups, marked by octahedra and tetrahedra, and clustered into two groups, marked by light and dark coloring.

association between groups and clusters is considered. The clustering itself is completely blind to treatment groups.

Further, the use of *nonmetric* clustering is not essential to this stage of the analysis. The association between clusters and treatment groups could be carried out after any clustering methodology, such as hierarchical or k-means clustering.

Now that each point has both a cluster and a group, the association between clusters and groups can be evaluated in a contingency table format. For instance, the points in Figure 10 would fill out the following table:

	Tetrahedra	Octahedra
Light	1	4
Dark	5	2

Under the null hypothesis that the treatment group has no effect on the data, the points in one treatment group would be just as likely to be in one cluster as another, and a uniform distribution of points in the contingency table would be expected. The Pearson χ^2 for the table can then be computed (Fienberg, 1985), to judge the significance of the effect (*i.e.* the probability, under the null hypothesis, of obtaining a χ^2 value at least as large as that of the observed table). Using this table, we get a value for χ^2 as follows:

$$\begin{aligned}\chi^2 &= \sum_{i,j} \frac{(N_{ij} - n_{ij})^2}{n_{ij}} \\ &= \frac{(1-2.5)^2}{2.5} + \frac{(4-2.5)^2}{2.5} + \frac{(5-3.5)^2}{3.5} + \frac{(2-3.5)^2}{3.5} \\ &\approx 3.09\end{aligned}$$

where N_{ij} is the actual cell count and n_{ij} is the expected cell count. With one degree of freedom (for a 2 × 2 table), this value can be looked up in a table of χ^2 probabilities to tell us that this (hypothetical) experiment shows a significant effect at the 90% level, but not the 95% level. Alternatively, a randomization or permutation test could be used to judge significance (Noreen, 1989).

Much toxicological testing uses four treatment groups, rather than two, but the strategy is the same. The data are clustered into four clusters, a (4 × 4) contingency table of treatment groups *vs.* clusters is assembled, and the significance of the effect is measured from the contingency table.

The group-cluster contingency table, however, can be used for more than simple hypothesis testing. The contents of the table can be examined to determine whether, for example, all four treatment groups were distinct, or only one or two of them. More sensitive measures of association in contingency tables, such as Guttman's λ or entropy (Goodman and Kruskal, 1954; Goodman and Kruskal, 1959; Goodman and Kruskal, 1963; Goodman and Kruskal, 1972) could also be used, to judge not *whether* an effect occurred, but *how strong* the effect was.

Nonmetric clustering and association analysis (NCAA) is a tool for evaluating the effects of treatments on multivariate systems. Because it is based on nonmetric clustering, the tool can be used on "messy" data, with missing points or with variates that do not obey assumptions of normality and homoscedasticity. In addition to evaluating the strength of the effect, the tool also provides insight into which of the variates are most strongly associated with the effect. It is also a "blind" test, in that the clustering is done independently of the treatment groups. This is useful in screening experiments for unlooked-for effects, such as edge-effects in mesocosms.

3 Projections

Nonmetric clustering can give some help in determining appropriate projections—the variables or parameters that are the most associated with the clustering are obvious candidates for a projection. However, if there are more than two or three of them, we have a (reduced) version of the same problem. As a result, some of linear projections, such as PCA or COA might be useful as a further insight into the nature of the patterns in the data. Each of these methods is actually a version of "projection pursuit", in its full generality: seek a projection of the data that maximizes some property of the data. PCA, for example, maximizes covariance or correlation.

Presently, we are working on a version of projection pursuit that maximizes the nonmetric associations we have seen, above. Instead of looking at the scatterplot matrix, projections onto all the original axes, and measuring the association in the quantile quadrats, we are working on an algorithm that will look at the association for quantile quadrats in an arbitrary projection. There is little mathematical theory to guide such a search, so it necessarily has to be heuristic. However, we have some promising early results which show that a good projection can be found reliably in reasonable time. Such a projection would be an adjunct to the standard projections, and reveal different patterns in the data.

4 Time

4.1 Spacetime Worms

A final problem confronting long-term testing is the integration of time into the analysis. Observations taken on the same system over a period of time are obviously correlated, so the analyst has the choice of investigating each day individually, and then combining the analyses, or analyzing all of the days together, but taking care that the time-correlations are considered. Time-series analysis is little help, because it is almost exclusively concerned with univariate changes over time—cycles, trends, *etc.* With a multivariate system changing over time, there is no such thing as going "up" or "down", there is only "hither" and "yon". There are a great many directions to go in 10-dimensional space.

One approach to understanding how systems evolve in time is to use some kind of one or two-dimensional projection, such as PCA or COA, and then examine the changes in the response "area" over time. In Section 5, the response of the SAM microcosms for copper sulfate, Jet-A, and JP-4, are plotted. The projections were PCA projections performed on the covariance matrix of the centered data (Pielou, 1984). A single projection of all data, for all sampling dates, was performed to get the two-dimensional points, but each plot only plots a single day's data. This way, the coordinates are consistent from day to day. The plots each consist of numbers, 1 through 4, showing the actual samples from treatment groups 1 to 4, and a circle drawn around the mean of each group. The radius of the circle is proportional to the average distance between points within the group. Thus, both the central trend of a group and its within-group variance can be graphically represented by the circle.

A better way of visualizing this day-to-day change in a projection of the data, however, is with a three-dimensional, interactive computer animation of the resulting spacetime "worm": the cylindrical surface generated by the response area circles. We have

implemented such a tool, and some examples (unfortunately, not in color) of what the same data look like are shown in Figures 11, 12, and 13.

4.2 Discrete Velocity, Curvature, and Torsion

In addition to visualization, however, we need to find some analysis tools that do not rely on human judgement, and which can be used as an aid in determining what to look at in the first place. While univariate time-series analysis does not help, it may be that some tools that have proven useful in the development of differential geometry will. Differential geometry is the study of motion and change in arbitrary spaces. Normally, this implies the use of calculus, and continuous real number fields. However, in our case, we rarely have the luxury of continuous monitoring, and must make due with a (small) set of discrete sampling points. Nevertheless, as outlined in Section 4.2.1, some of the concepts can be generalized to discrete-step processes. It is possible that n-dimensional velocity, acceleration, curvature and torsion could provide conceptual handles on the nature of changes in high-dimensioned space.

Intuitively, velocity, curvature and torsion can be understood by considering the motion of the earth through space. A tangent to the circle the earth follows around the earth points in the direction of our "velocity vector". A line from the earth toward the sun points in the direction of our "curvature vector". Now, the whole solar system, sun, planets and all, is travelling roughly in the direction of the star Vega; thus, in addition to going around and around in a two-dimensional plane, we are also spiralling out of that plane in the direction of Vega. Our "torsion vector" points toward Vega. In other words, the torsion vector points in the direction the corkscrew moves when it is twisted into the cork. Figure 14 shows how these vectors look on a spiral.

Some mathematical definitions of these quantities follow. We have yet to determine whether they will be of value in assessing impacts over time.

4.2.1 Mathematical Definitions

A *finite, discrete, parameterized curve* (fdp-curve) in n-space is a pair, $\langle T, s \rangle$, where T is a finite set of real numbers $T = \{t_1 \ldots t_m\}$ (which we assume, for convenience, are in order: $t_i < t_j \iff i < j$), and s is a function $s : T \rightarrow \mathbf{R}^n$. Intuitively, T represents the times at which the system was sampled, and $s(t)$ represents the state of the system at time t.

Given an fdp-curve, $\langle T, s \rangle$, we can define the (discrete)
velocity vector V of s at time $t_i \in T$ as the change per unit time:

$$V(t_i) = (s(t_{i+1}) - s(t_i))/(t_{i+1} - t_i)$$

for $i \leq i \leq m - 1$. The *velocity* is the length of the velocity vector:

$$v(t_i) = \| V(t_i) \|$$

When the appropriate velocities are nonzero, the *curvature vector* C is the change in the velocity vector, after normalizing to remove linear acceleration:

$$C(t_i) = \left(\frac{V(t_{i+1})}{\| V(t_{i+1}) \|} - \frac{V(t_i)}{\| V(t_i) \|} \right) / (t_{i+1} - t_i)$$

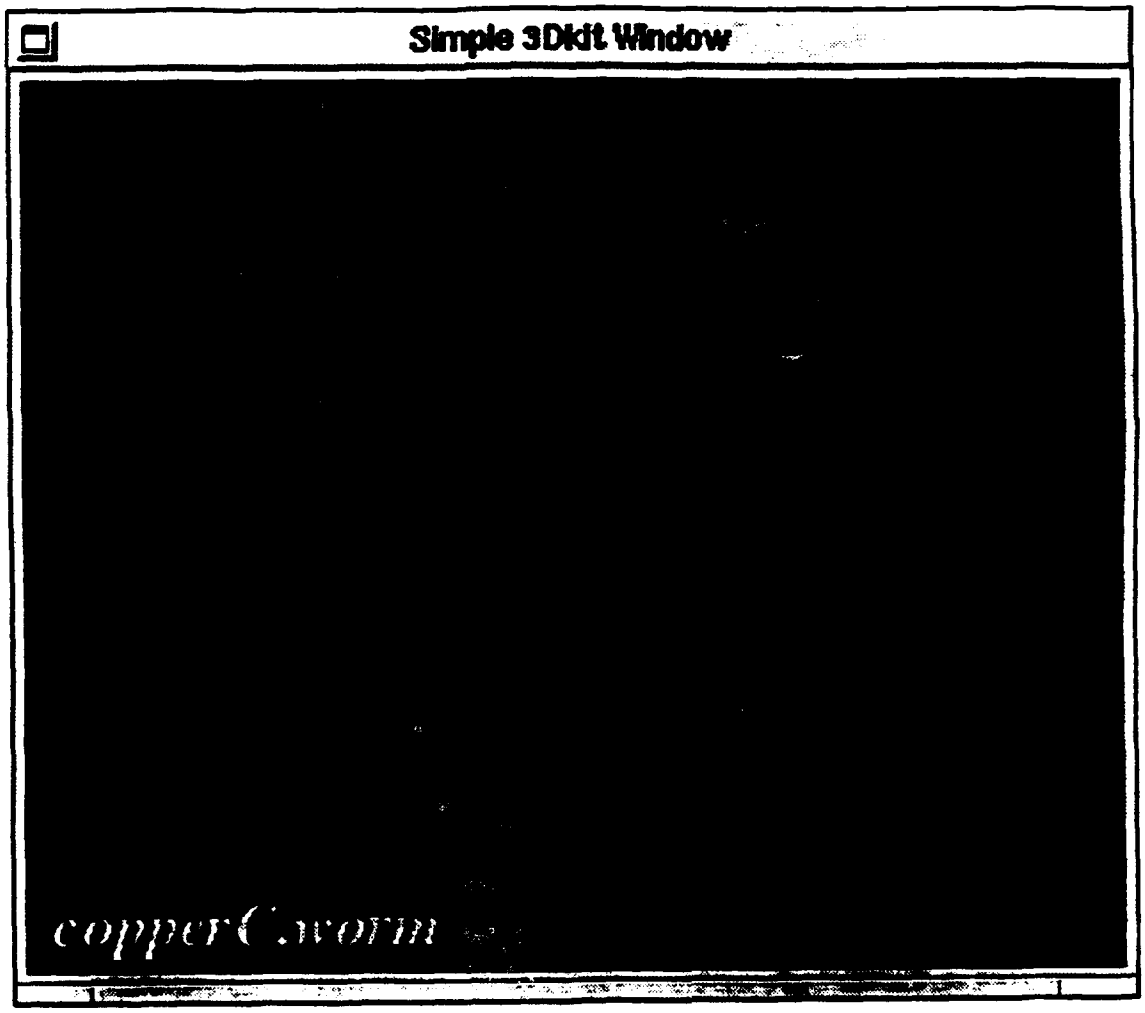


Figure 11: Spacetime worms of the PCA projections of the copper sulfate SAM response areas for treatment groups 1 and 4.

Figure 12: Spacetime worms of the PCA projections of the Jet-A SAM response areas for treatment groups 1 and 4.

Figure 13: Spacetime worms of the PCA projections of the JP-4 SAM response areas for treatment groups 1 and 4.

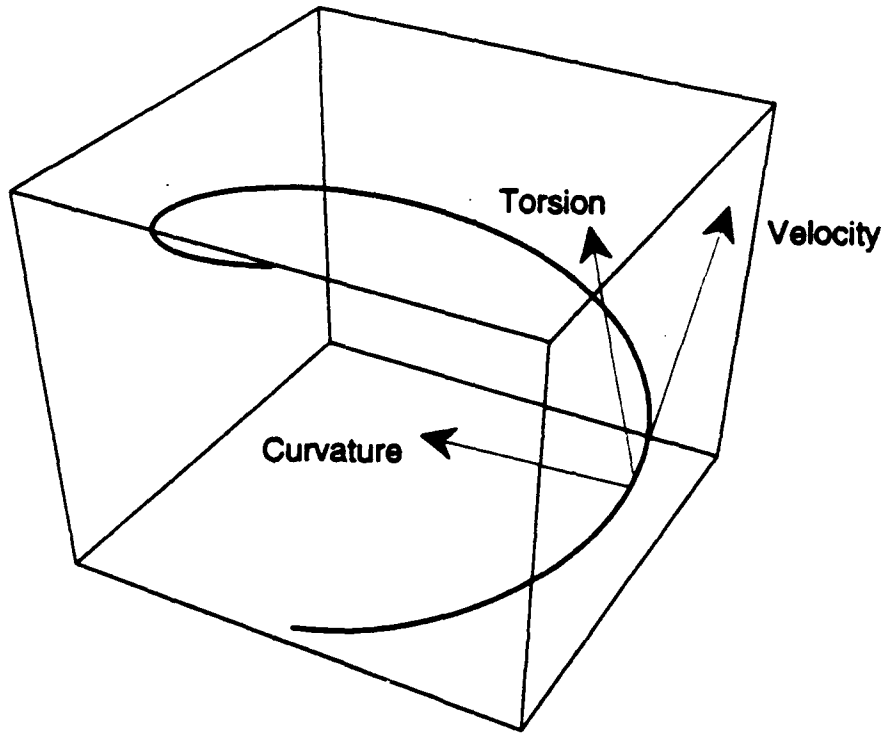


Figure 14: The velocity, curvature, and torsion vectors at a particular point along a spiral.

for $1 \leq i \leq m-2$. The *curvature* is the length of the curvature vector:

$$c(t_i) = \| C(t_i) \|$$

The *torsion vector* and *torsion* are defined by analogy with the Frenet formulas (O'Neill, 1966), so as to be nearly perpendicular to the velocity and curvature vectors when they are not changing too rapidly:

$$\begin{aligned} T(t_i) &= (C(t_{i+1}) - C(t_i))/(t_{i+1} - t_i) + c(t_i)V(t_i)/ \| V(t_i) \| \\ t(t_i) &= \| T(t_i) \| \end{aligned}$$

It should be remarked that velocity at time t_i requires the values of t_i and t_{i+1}, curvature requires t_i, t_{i+1}, and t_{i+2}, and torsion requires t_i, t_{i+1}, t_{i+2}, and t_{i+3}. Thus, if the original curve has m points, only $m-3$ points will have velocity, curvature, and torsion all defined.

As $t_{i+1} \rightarrow t_i$ these vectors become the tangent, normal and binormal vectors of a continuous curve in n-space.

References

Asimov, D. (1985). The grand tour. *SIAM Journal of Scientific and Statistical Computing*, 6:128–143.

Fienberg, S. E. (1985). *The Analysis of Cross-Classified Categorical Data*. MIT Press, Cambridge, MA.

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49:732–764.

Goodman, L. A. and Kruskal, W. H. (1959). Measures of association for cross classifications ii: further discussion and references. *Journal of the American Statistical Association*, 54:123–163.

Goodman, L. A. and Kruskal, W. H. (1963). Measures of association for cross classifications iii: approximate sampling theory. *Journal of the American Statistical Association*, 58:310–364.

Goodman, L. A. and Kruskal, W. H. (1972). Measures of association for cross classifications iv: simplification of asymptotic variances. *Journal of the American Statistical Association*, 67:415–421.

Huber, P. J. (1985). Projection pursuit. *Annals of Statistics*, 13:435–475.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ.

Matthews, G. and Hearne, J. (1991). Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175–184.

Matthews, G. B., Matthews, R. A., and Hachmöller, B. (1991a). Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(11):2184–2190.

Matthews, R. A., Matthews, G. B., and Ehinger, W. J. (1991b). Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modelling*, 53:167-187.

Noreen, E. W. (1989). *Computer Intensive Methods for Testing Hypotheses*. Wiley-Interscience, New York, NY.

O'Neill, B. (1966). *Elementary Differential Geometry*. Academic Press, New York.

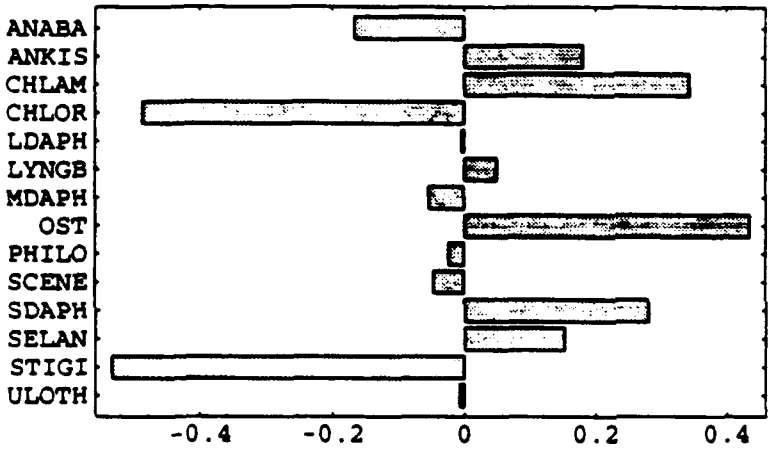
Pielou, E. C. (1984). *The Interpretation of Ecological Data. A Primer on Classification and Ordination*. Wiley-Interscience, New York, NY.

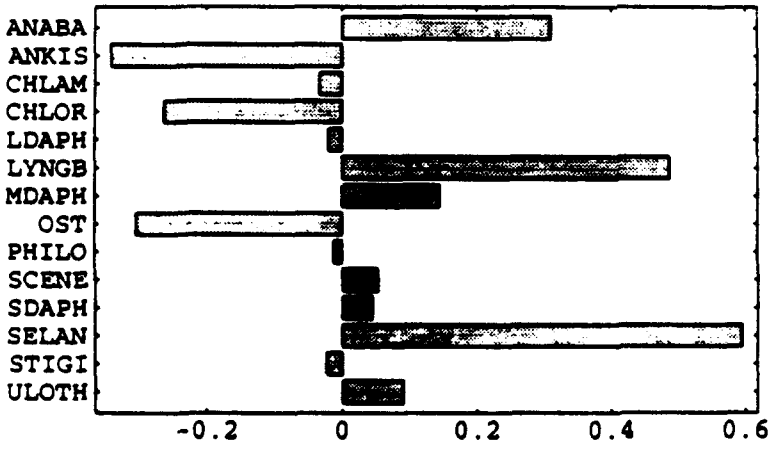
Smith, E. P., Pontasch, K. W., and Cairns, Jr., J. (1990). Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.*, 24(4):507-514.

5 Response area plots

Copper PCA

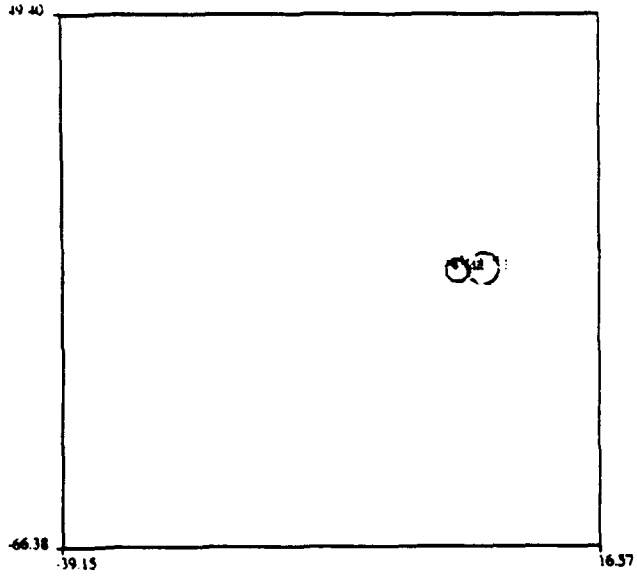
	First PCA	Second PCA
ANABA	-0.1666	0.3100
ANKIS	0.1814	-0.3410
CHLAM	0.3438	-0.0333
CHLOR	-0.4867	-0.2626
LDAPH	-0.0042	-0.0198
LYNGB	0.0504	0.4862
MDAPH	-0.0540	0.1442
OST	0.4350	-0.3038
PHILO	-0.0238	-0.0116
SCENE	-0.0466	0.0519
SDAPH	0.2804	0.0443
SELAN	0.1525	0.5949
STIGI	-0.5335	-0.0226
ULOTH	-0.0071	0.0902





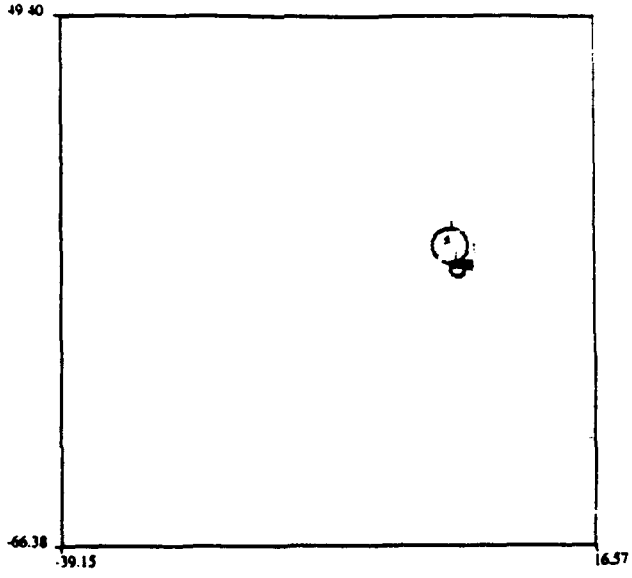
Copper, Day 11

PCA.
First component horizontal, second component vertical.



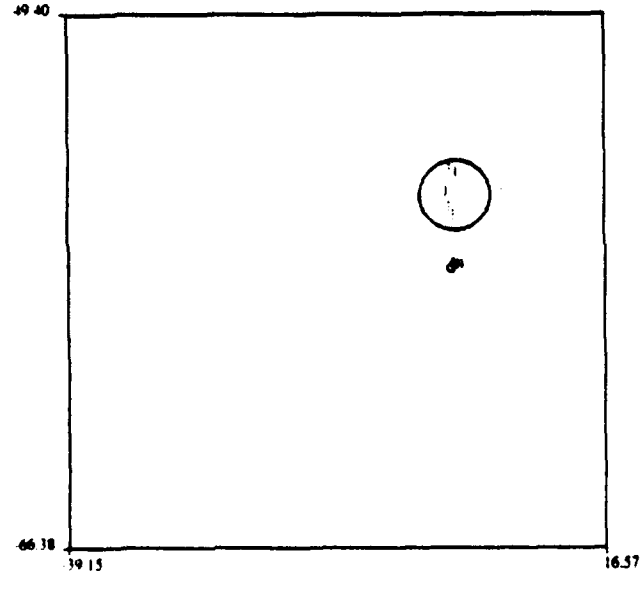
Copper, Day 14

PCA.
First component horizontal, second component vertical.



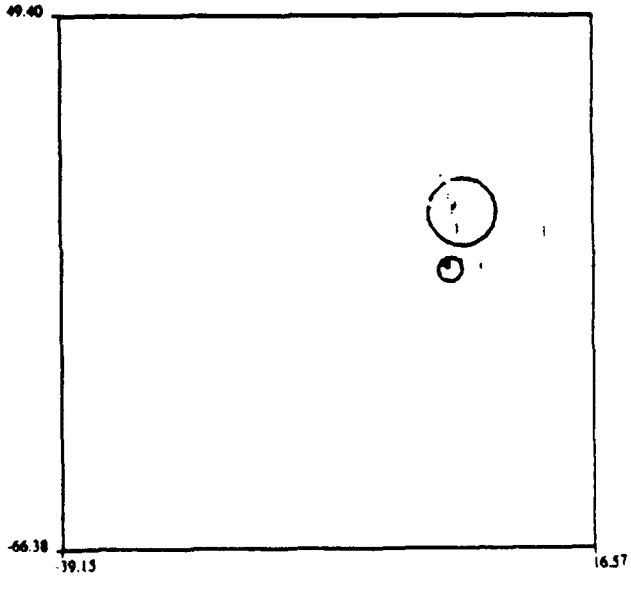
Copper, Day 18

PCA.
First component horizontal, second component vertical.



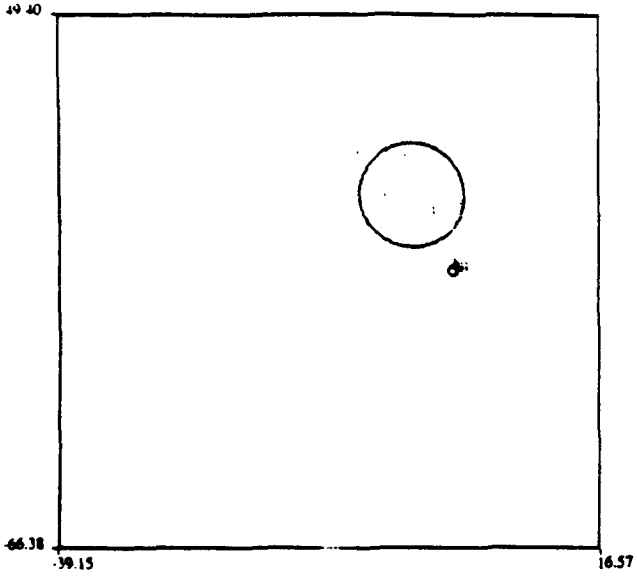
Copper, Day 21

PCA.
First component horizontal, second component vertical.



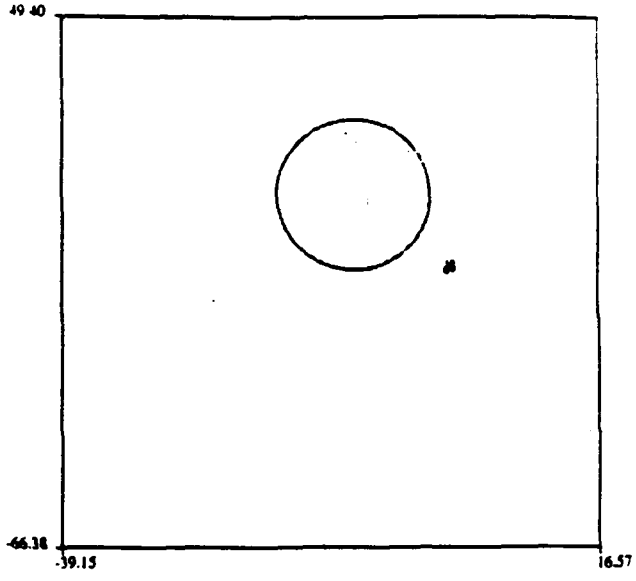
Copper, Day 25

PCA.
First component horizontal, second component vertical.



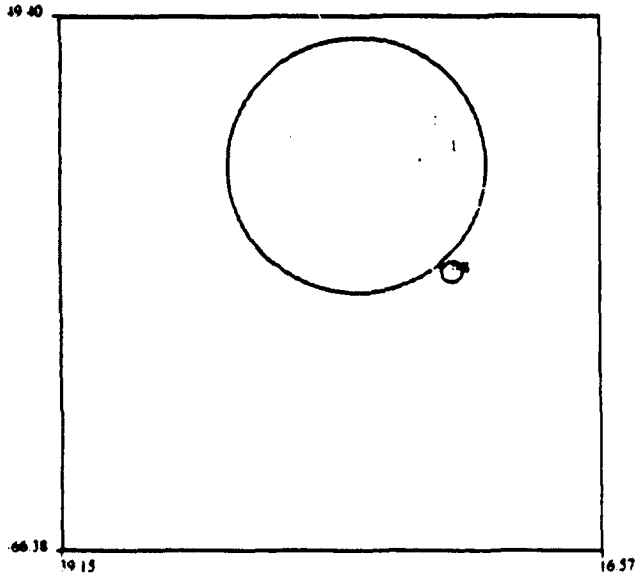
Copper, Day 28

PCA.
First component horizontal, second component vertical.



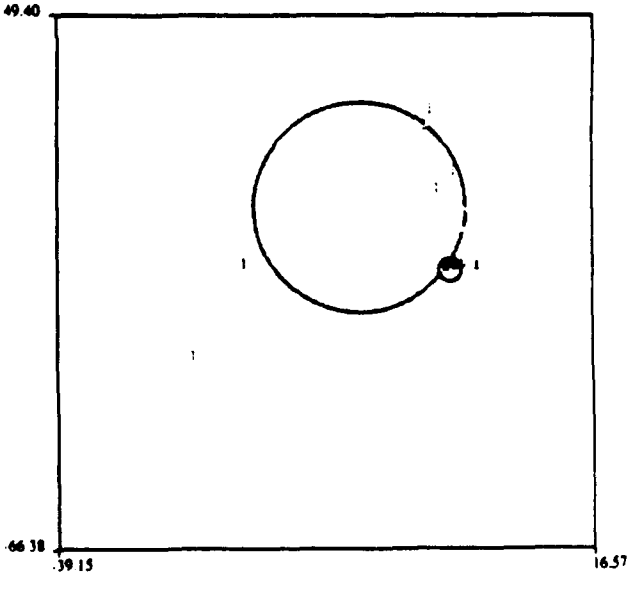
Copper, Day 32

PCA.
First component horizontal, second component vertical.



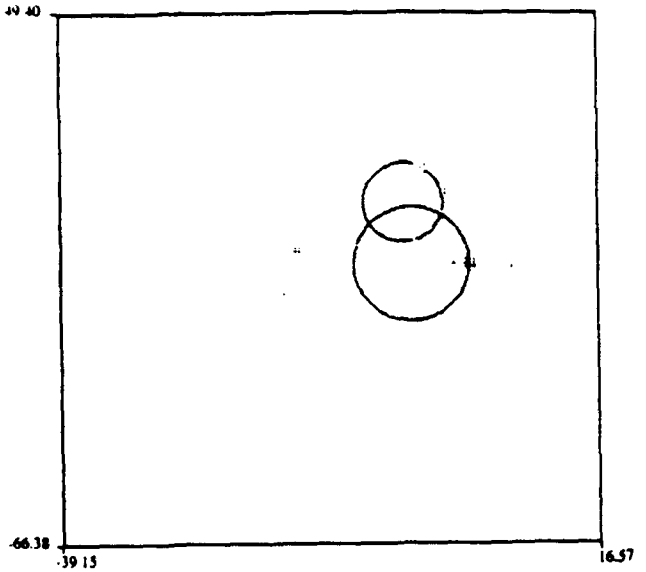
Copper, Day 35

PCA.
First component horizontal, second component vertical.



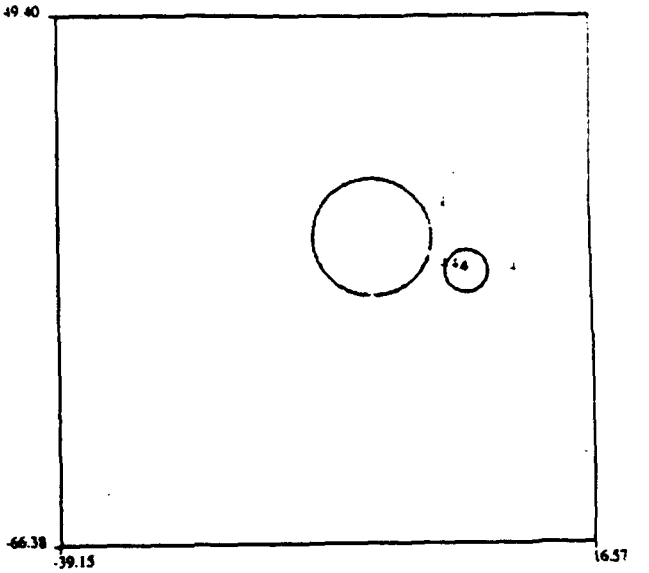
Copper, Day 39

PCA.
First component horizontal, second component vertical.



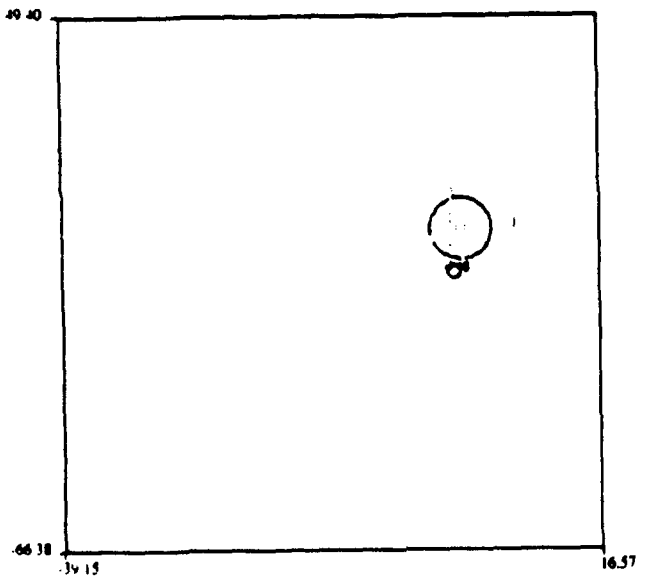
Copper, Day 42

PCA.
First component horizontal, second component vertical.



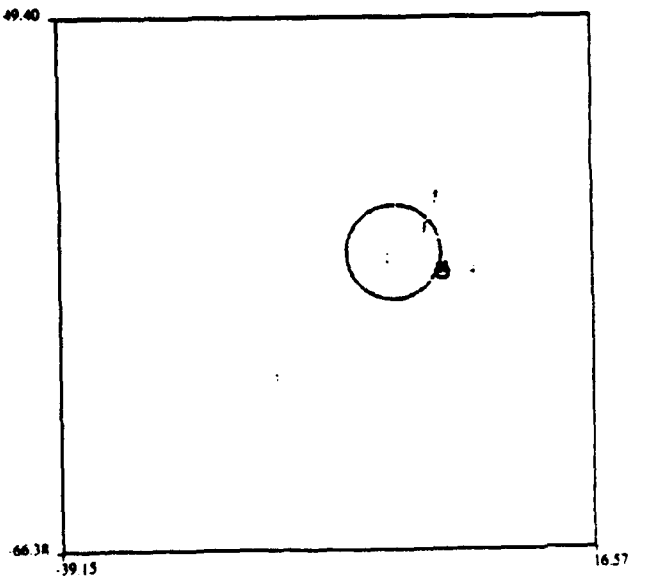
Copper, Day 46

PCA.
First component horizontal, second component vertical.



Copper, Day 49

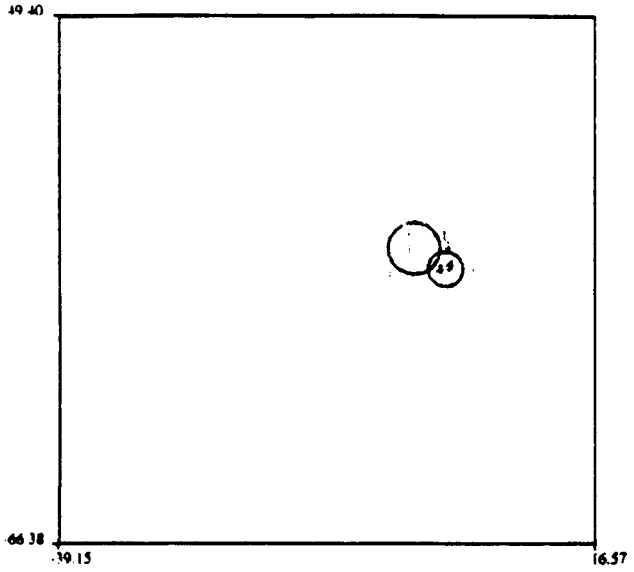
PCA.
First component horizontal, second component vertical.



Copper, Day 53

PCA.

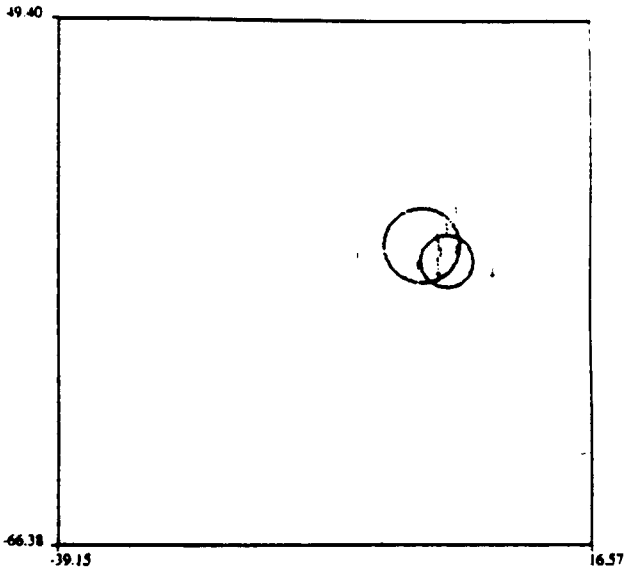
First component horiz al. second component vertical.



Copper, Day 56

PCA.

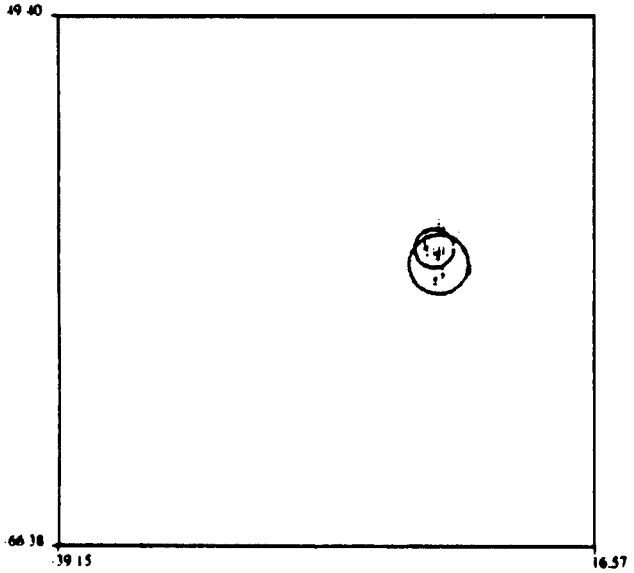
First component horizontal. second component vertical.



Copper, Day 60

PCA.

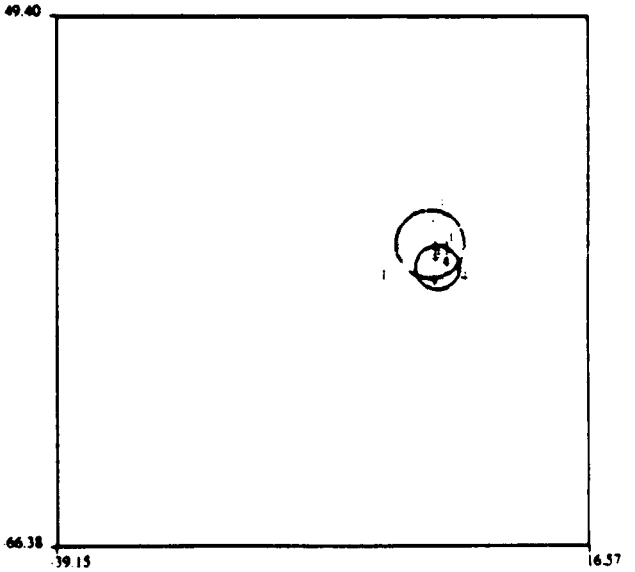
First component horizontal. second component vertical.



Copper, Day 63

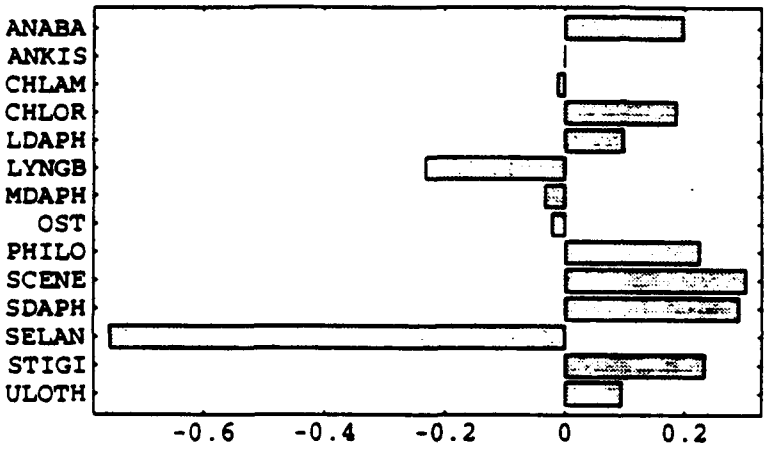
PCA.

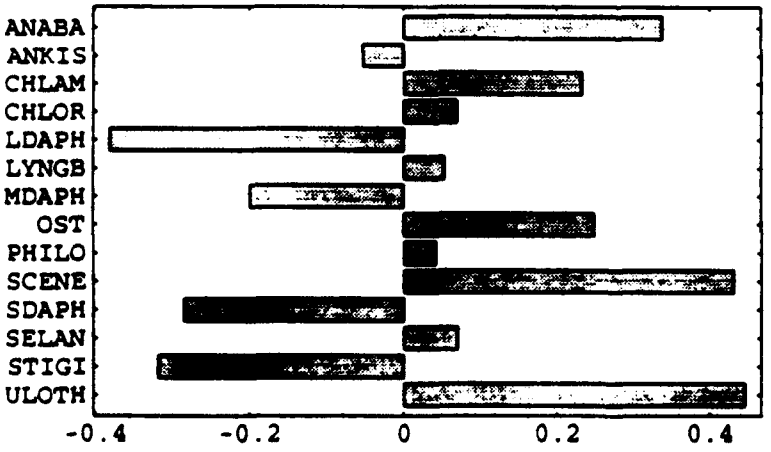
First component horizontal. second component vertical.



JetA PCA

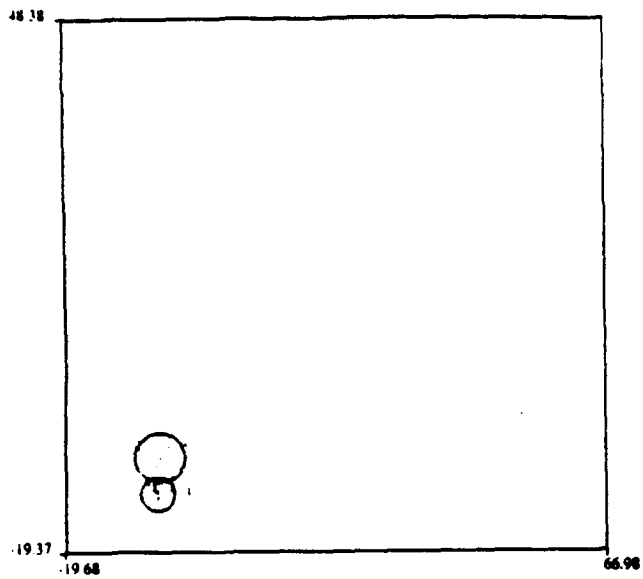
	First PCA	Second PCA
ANABA	0.1984	0.3382
ANKIS	-0.0003	-0.0534
CHLAM	-0.0120	0.2326
CHLOR	0.1868	0.0693
LDAPH	0.0982	-0.3811
LYNGB	-0.2328	0.0529
MDAPH	-0.0337	-0.1992
OST	-0.0230	0.2483
PHILO	0.2247	0.0419
SCENE	0.3016	0.4312
SDAPH	0.2902	-0.2847
SELAN	-0.7553	0.0705
STIGI	0.2338	-0.3179
ULOTH	0.0946	0.4465





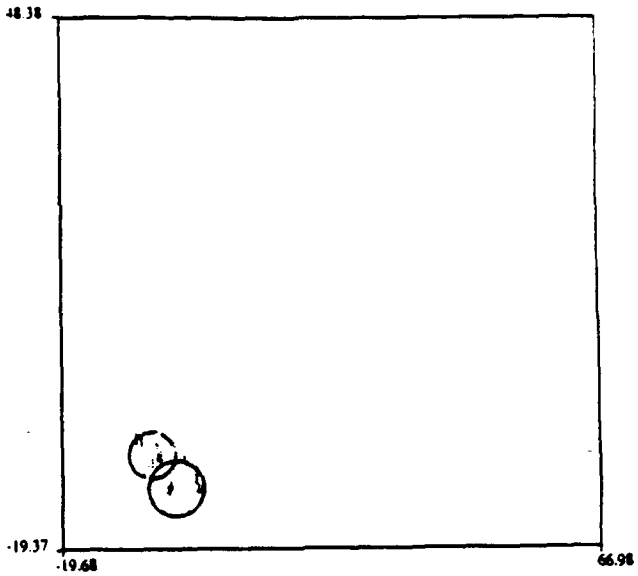
JetA. Day 11

PCA.
First component horizontal. second component vertical.



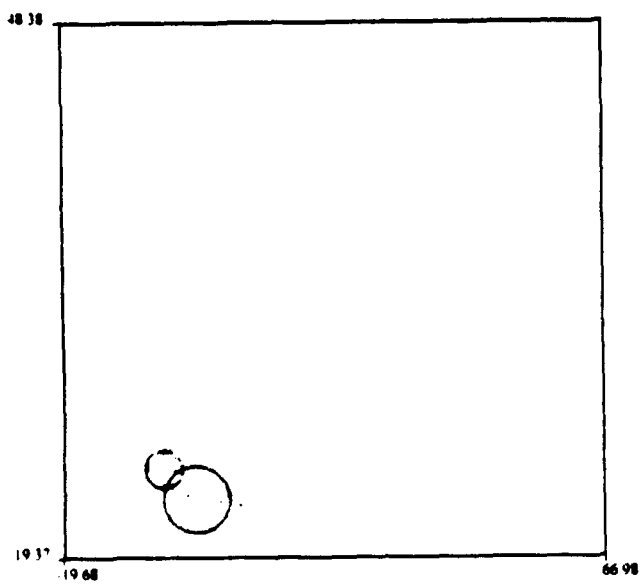
JetA, Day 14

PCA.
First component horizontal. second component vertical.



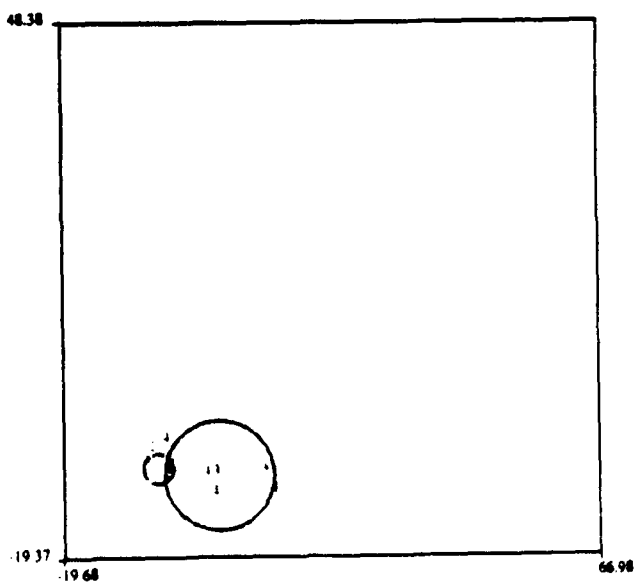
JetA. Day 18

PCA.
First component horizontal. second component vertical.



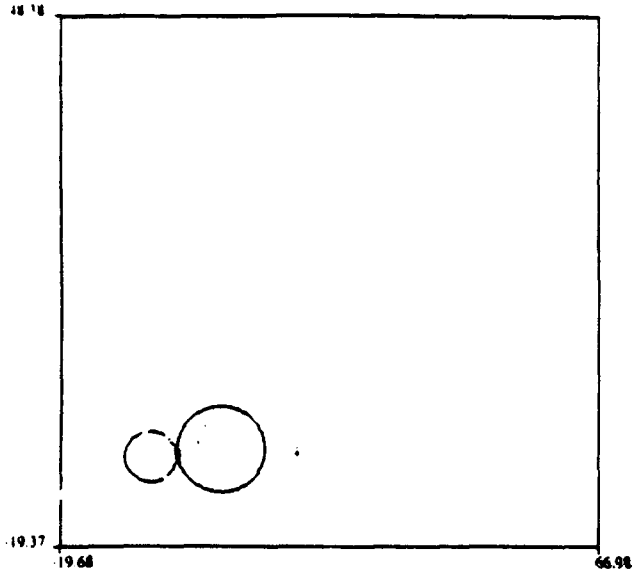
JetA, Day 21

PCA.
First component horizontal. second component vertical.



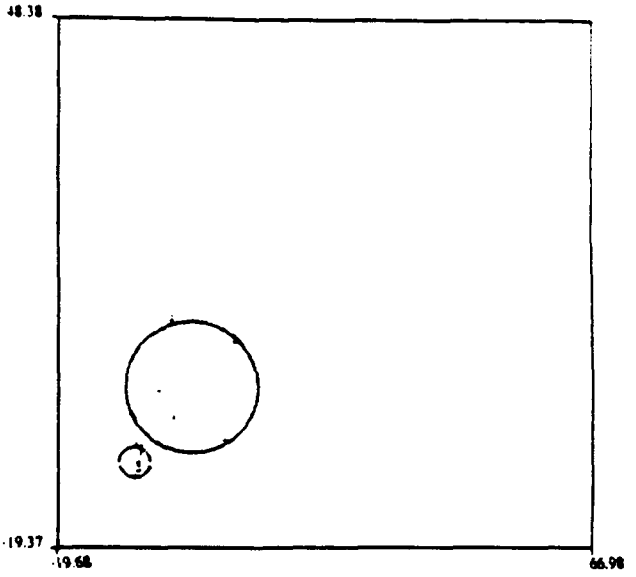
JetA, Day 25

PCA.
First component horizontal, second component vertical.



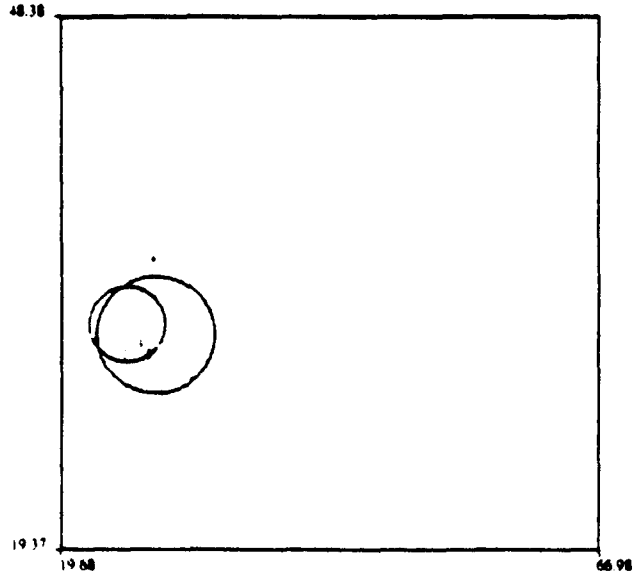
JetA, Day 28

PCA.
First component horizontal, second component vertical.



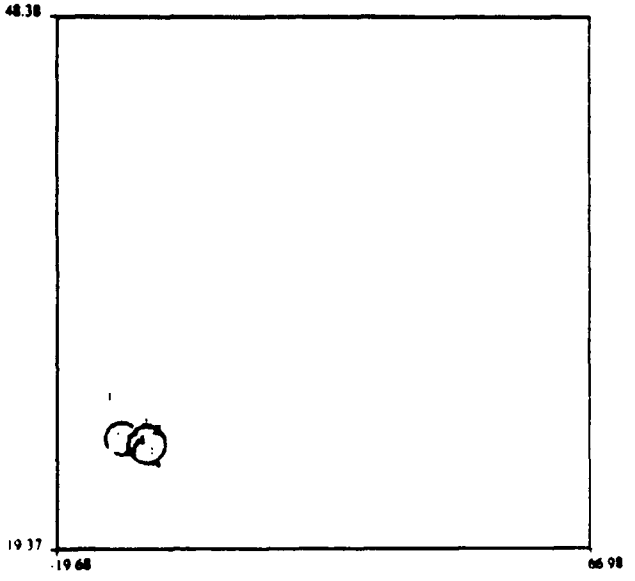
JetA, Day 32

PCA.
First component horizontal, second component vertical.



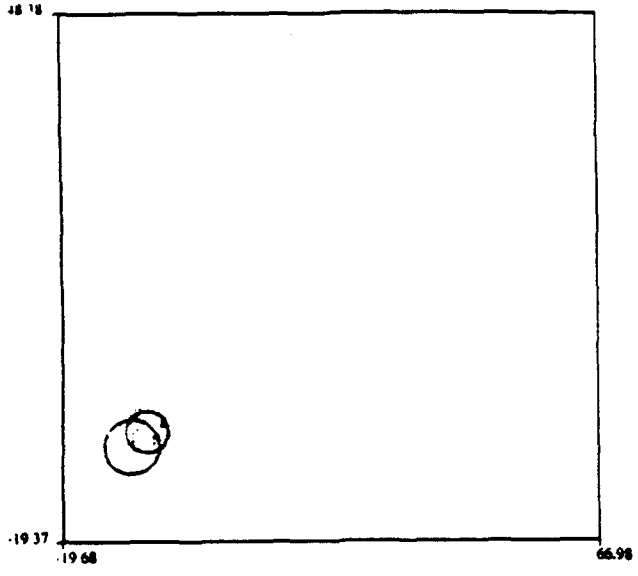
JetA, Day 35

PCA.
First component horizontal, second component vertical.



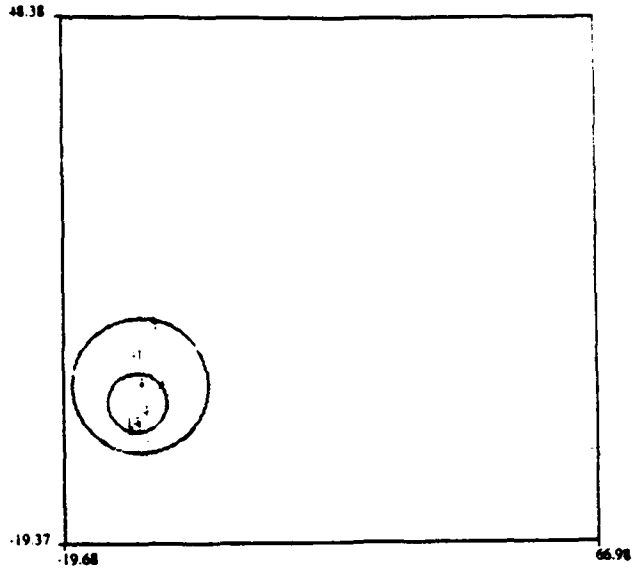
JetA, Day 39

PCA.
First component horizontal, second component vertical.



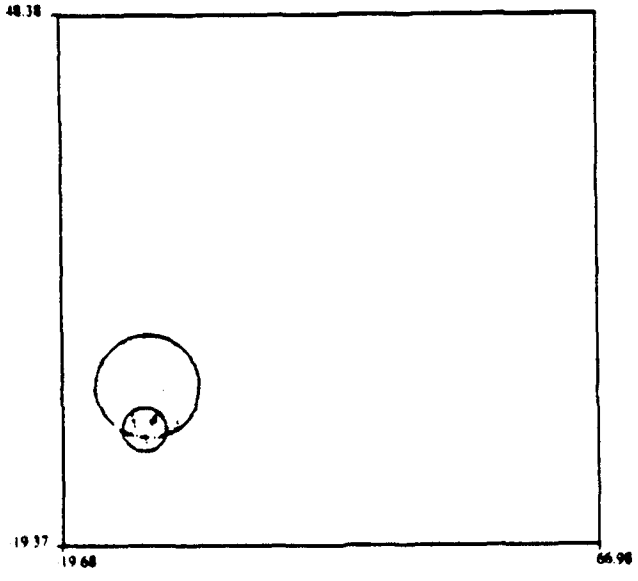
JetA, Day 42

PCA.
First component horizontal, second component vertical.



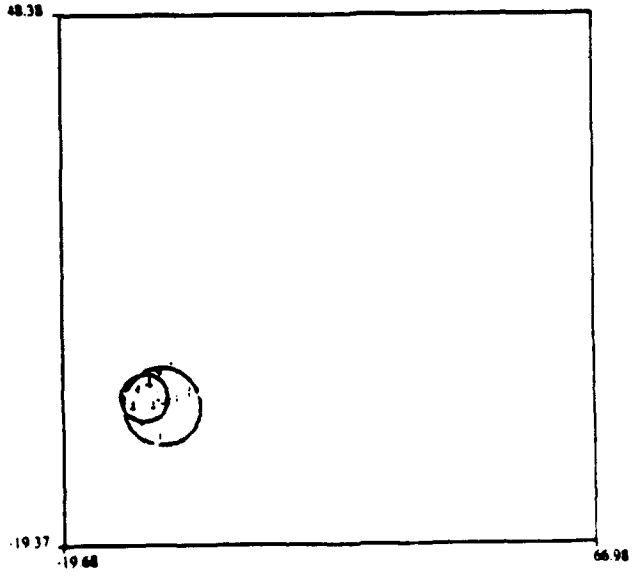
JetA, Day 46

PCA.
First component horizontal, second component vertical.



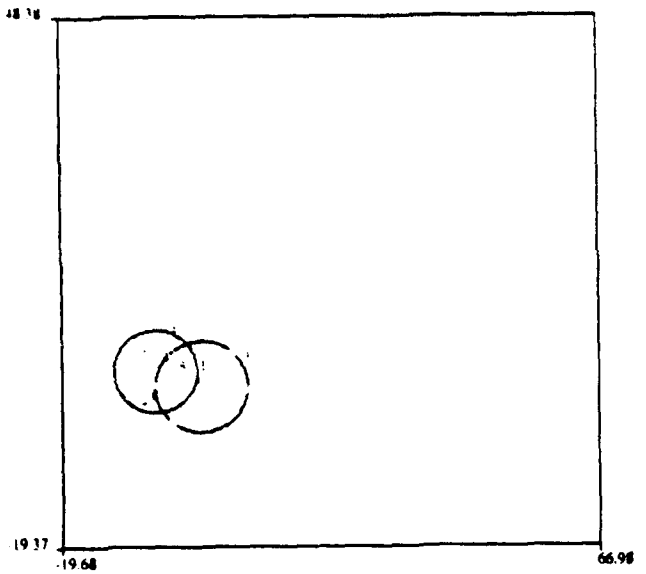
JetA, Day 49

PCA.
First component horizontal, second component vertical.



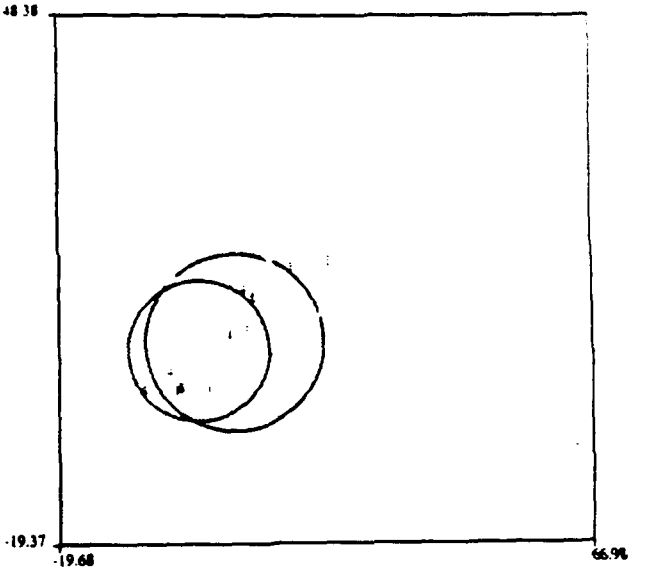
JetA, Day 53

PCA.
First component horizontal, second component vertical.



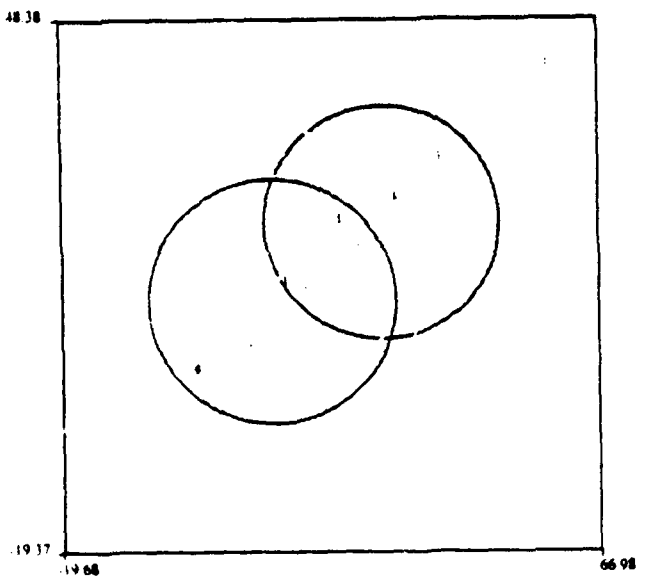
JetA, Day 56

PCA.
First component horizontal, second component vertical.



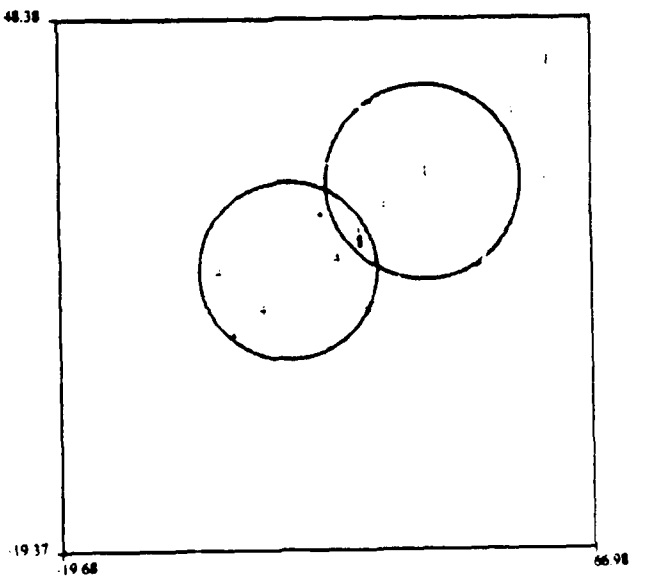
JetA, Day 60

PCA.
First component horizontal, second component vertical.



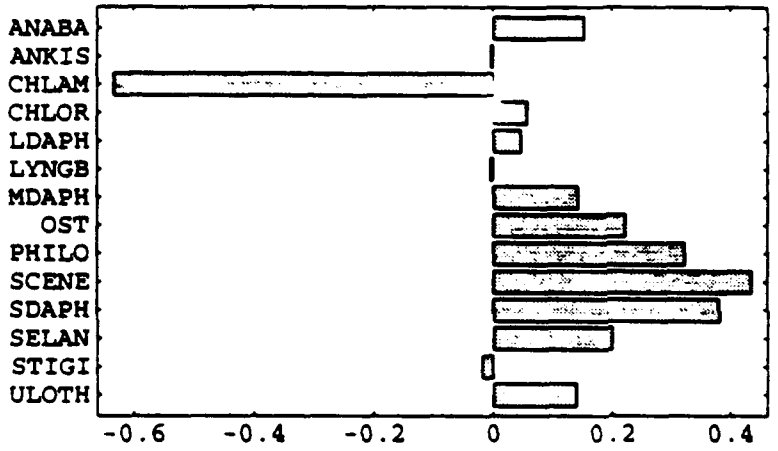
JetA, Day 63

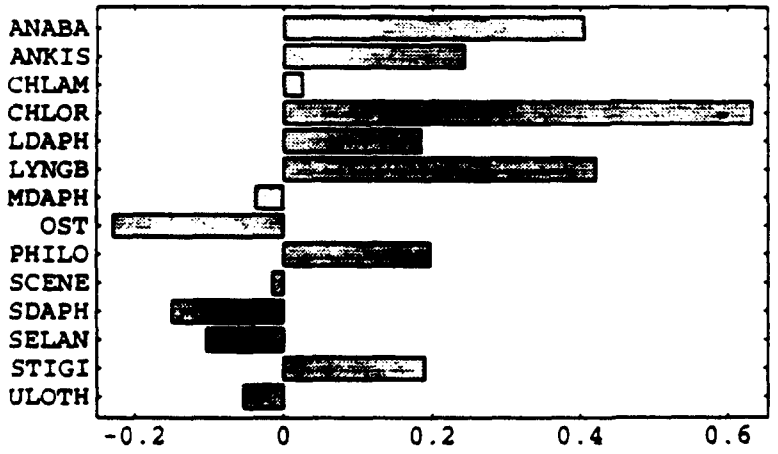
PCA.
First component horizontal, second component vertical.



JP4 PCA

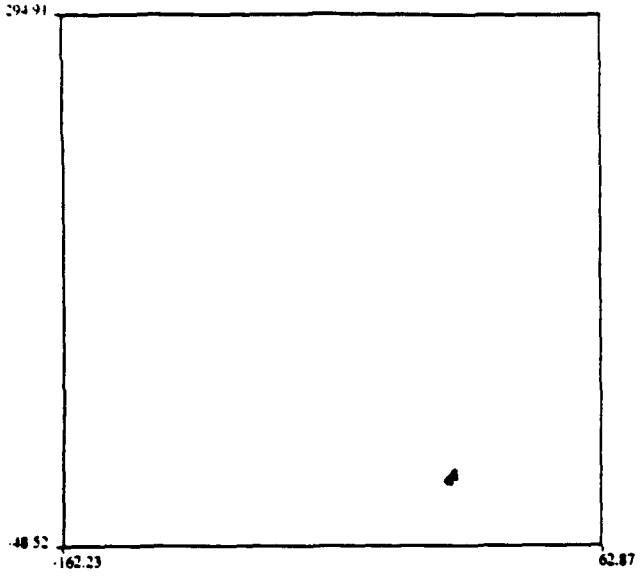
	First PCA	Second PCA
ANABA	0.1530	0.4048
ANKIS	-0.0043	0.2430
CHLAM	-0.6340	0.0250
CHLOR	0.0578	0.6327
LDAPH	0.0472	0.1842
LYNGB	-0.0049	0.4209
MDAPH	0.1428	-0.0372
OST	0.2228	-0.2288
PHILO	0.3225	0.1966
SCENE	0.4354	-0.0153
SDAPH	0.3811	-0.1492
SELAN	0.2001	-0.1035
STIGI	-0.0187	0.1906
ULOTH	0.1396	-0.0536





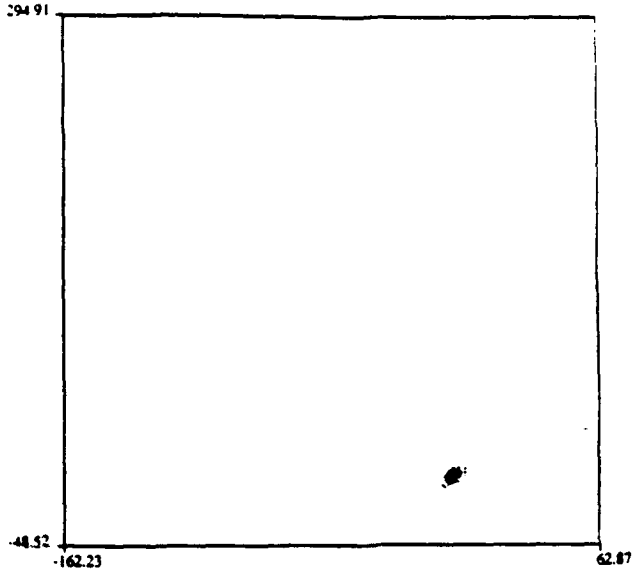
JP4, Day 11

PCA.
First component horizontal, second component vertical.



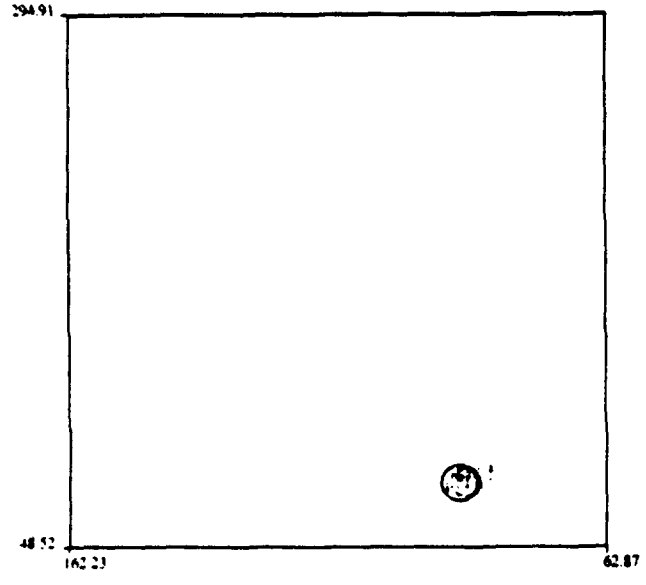
JP4, Day 14

PCA.
First component horizontal, second component vertical.



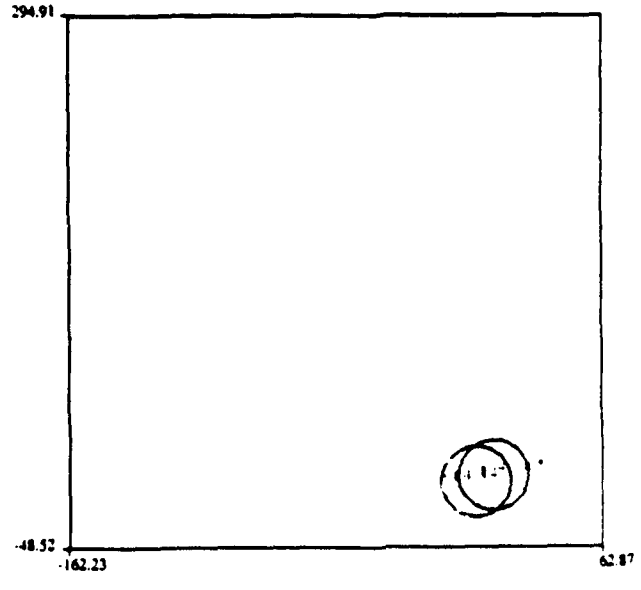
JP4, Day 18

PCA.
First component horizontal, second component vertical.



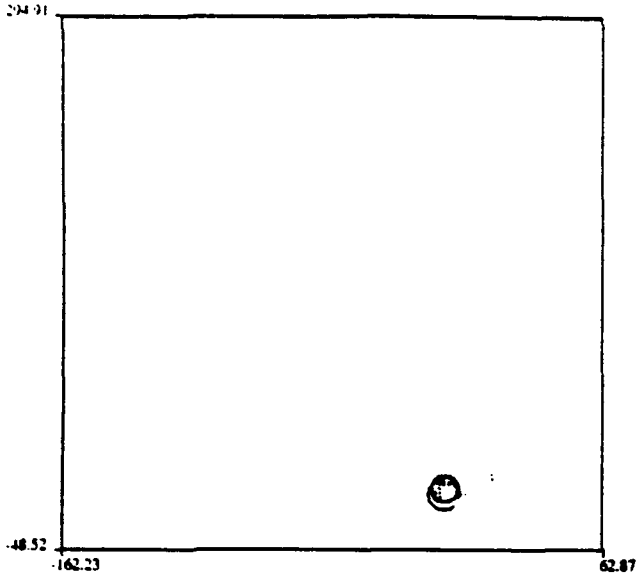
JP4, Day 21

PCA.
First component horizontal, second component vertical.



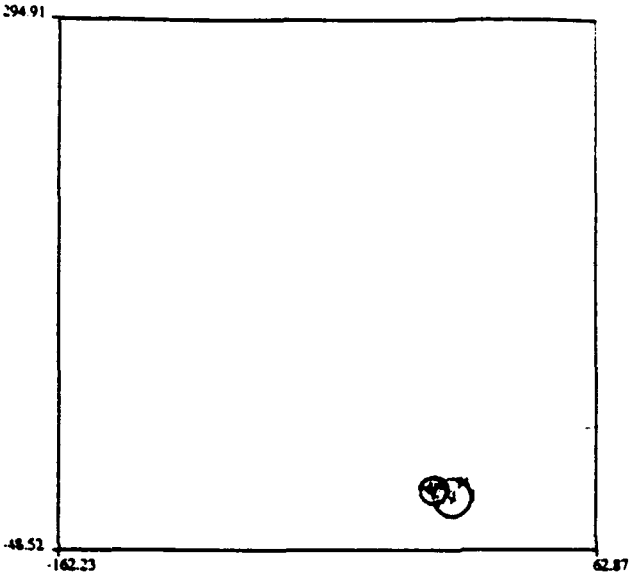
JP4, Day 25

PCA.
First component horizontal, second component vertical.



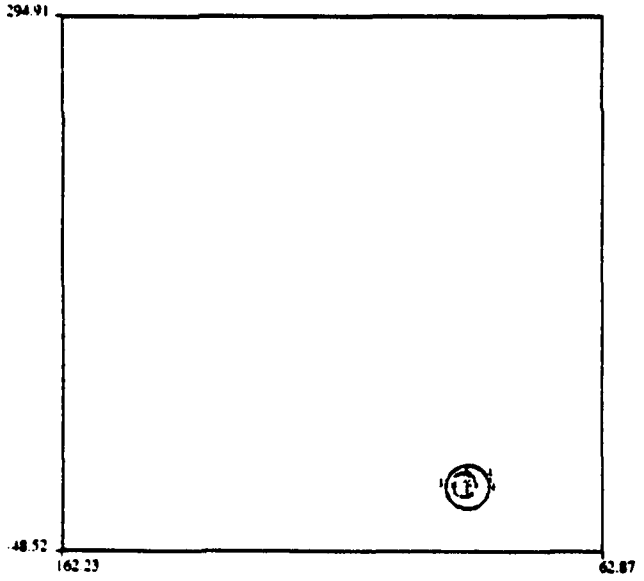
JP4, Day 28

PCA.
First component horizontal, second component vertical.



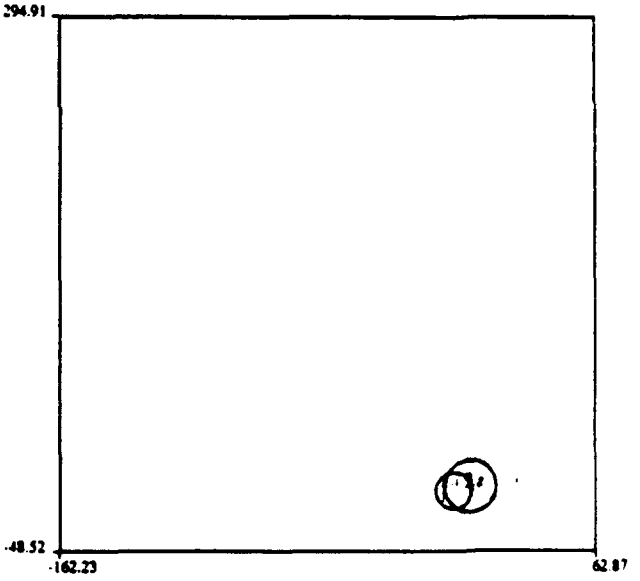
JP4, Day 32

PCA.
First component horizontal, second component vertical.



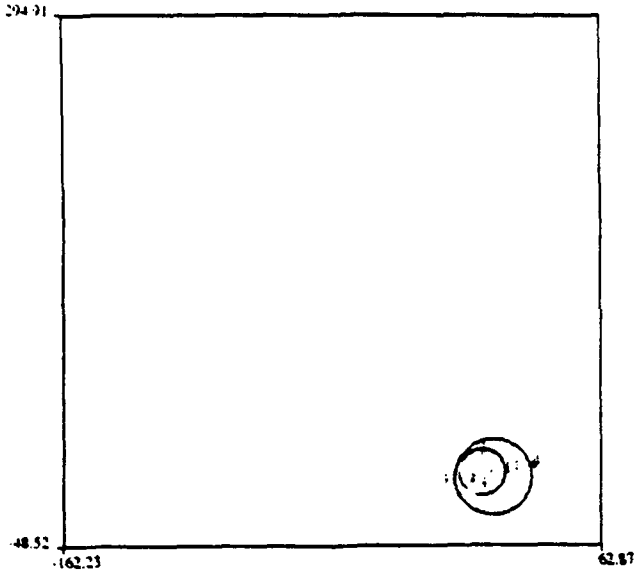
JP4, Day 35

PCA.
First component horizontal, second component vertical.



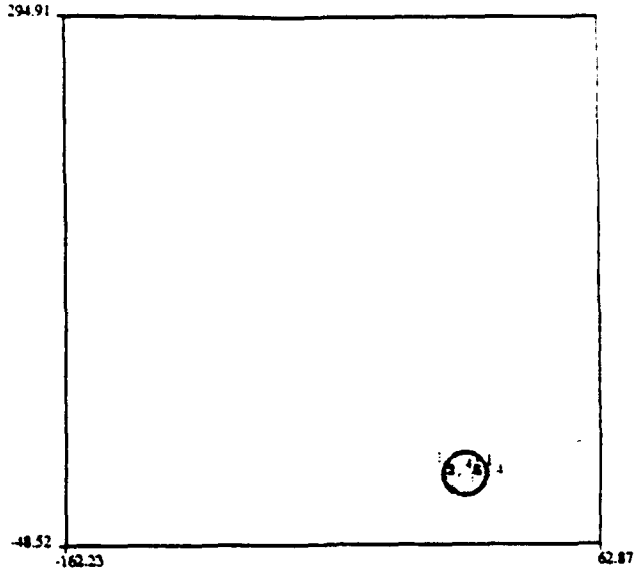
JP4, Day 39

PCA.
First component horizontal, second component vertical.



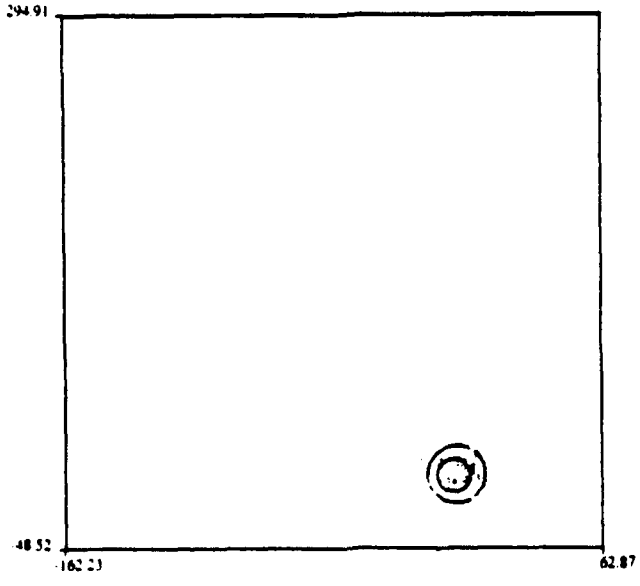
JP4, Day 42

PCA.
First component horizontal, second component vertical.



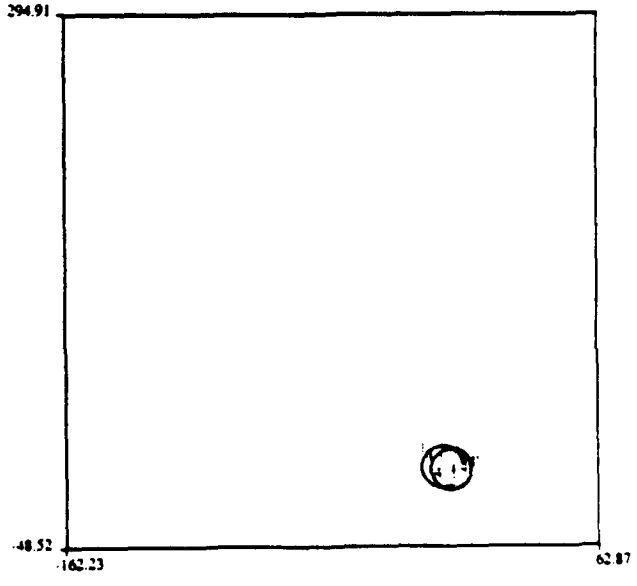
JP4, Day 46

PCA.
First component horizontal, second component vertical.



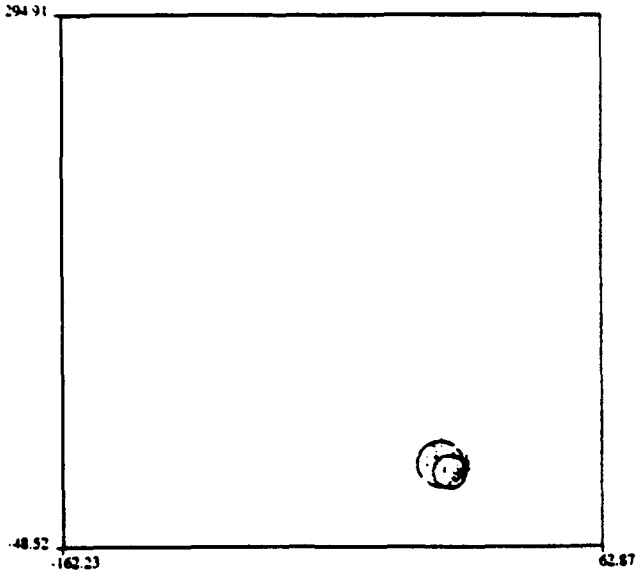
JP4, Day 49

PCA.
First component horizontal, second component vertical.



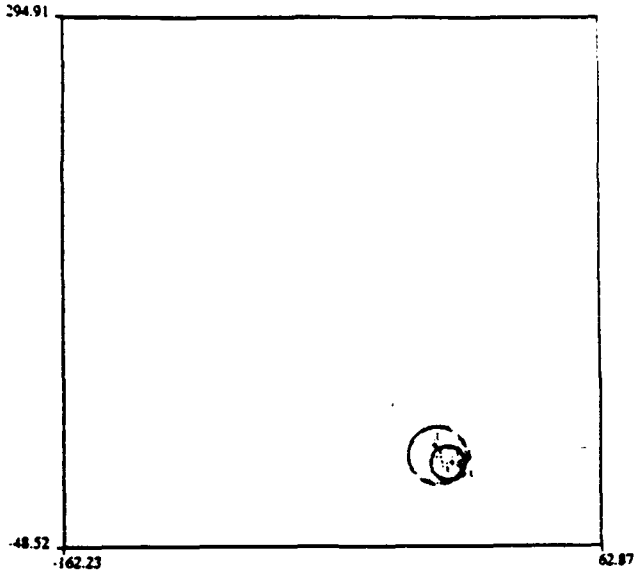
JP4, Day 53

PCA.
First component horizontal, second component vertical.



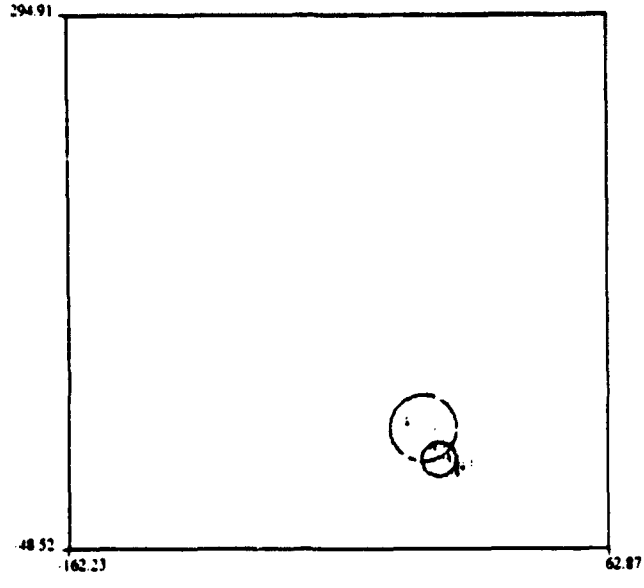
JP4, Day 56

PCA.
First component horizontal, second component vertical.



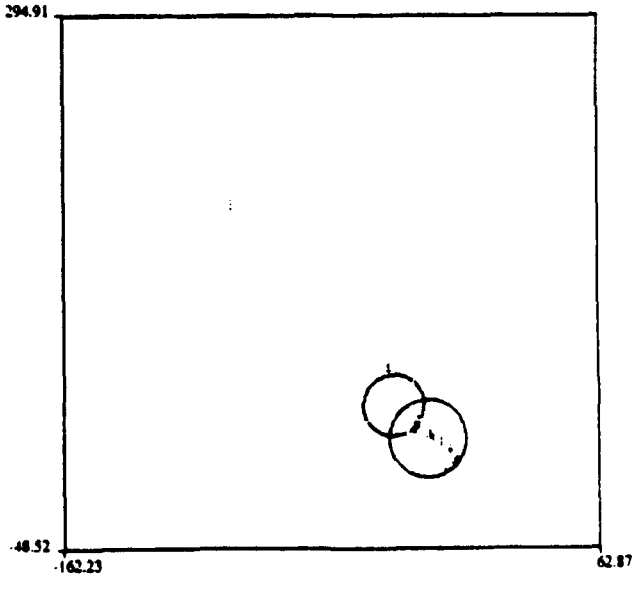
JP4, Day 60

PCA.
First component horizontal, second component vertical.



JP4, Day 63

PCA.
First component horizontal, second component vertical.



Technical Report: Development of a Nonmetric Clustering User Interface

Michael J. Roze
Computer Science Department
Western Washington University

April 21, 1993

Abstract

This paper describes a user-friendly frontend to the RIFFLE statistical clustering program [5]. I present an informal user's guide to the interface and discuss some planned extensions to the interface based on current research. The development of this interface makes a significant data analysis tool accessible to researchers of all disciplines. As work continues the research team is developing the interface as an implementation of a unified approach to the statistical analysis of similar datasets based on the experience gained in the current studies [6, 7, 8, 10].

Contents

List of Figures

1 Introduction

Clustering is a data analysis technique that attempts to fit the data to a number of clusters, or subpopulations, each with distinct properties. Clustering algorithms attempt to group the data points by maximizing within cluster similarity and simultaneously minimizing between cluster similarity. RIFFLE implements a new approach to **nonmetric clustering**, and was developed by Matthews and Hearne [5].

In RIFFLE'S original (command line) form the program requires the user to have a fairly thorough understanding of the input arguments and their effects. However, one of the goals of this research is to make this statistical clustering tool available to researchers in other disciplines. This motivated a project to develop a graphical user interface making the program easier to use, and giving graphical results in addition to the original text output.

The following sections will explain how to use the interface (in the process exploring some of its capabilities), and briefly describe what is planned for future development.

The interface is implemented on a NeXT computer.

2 Using the Interface

2.1 Data Files

Input files are expected to be real numbers or integers separated by white space (spaces or tabs). The data file should look something like this:

```
5.1   3.5   1.4   0.2
4.9   3.0   1.7   0.2
4.7   3.2   1.3   0.4
```

In this case the number of features would be four, unless there is a description file (discussed in Section 2.2) that indicates a different number.

The data file must be organized in a specific, but sensible, way. As an example we'll use Anderson's Iris data [1, 2, 3, 4]. In this dataset each data point would represent the measurements of four features (or attributes) on a single Iris flower: petal length, petal width, sepal length, and sepal width. Since 150 flowers were inspected in the study, there are 150 rows in the data file, and in each row there are four numbers, representing the values of each of the four features for that specific flower. Thus the data file is a 150 row by 4 column matrix of numbers. Each column of data must represent the

values measured for just one feature. That is, if the first number for the first flower is the petal length, then the first number for all the other rows must also be petal length, and so on for all the columns.

RIFFLE is able to accept data with missing values, however, the number of values in each row must be consistent over the entire file. Missing values are explicitly indicated by valu less than or equal to **-99**, in contrast the regular data values are expected to be **zero**, **positive integer**, or **positive real** numbers, but not negative numbers. If a data point is missing the x or y feature (or both) it will not be plotted. While RIFFLE can accommodate data that has some missing data values (**-99**), it probably does not make sense to run the interface with data that, for example, is missing half of the data points.

In addition, the association analysis (Section 2.6), and plot by treatment group (Section 2.7.4) features require the data file to be structured in a way that allows the interface to distinguish the treatment groups.

- All of the points in a group must be listed consecutively in the data file, and
- All groups must have the same number of points.

With three groups, the interface will consider the first third of the data points as group "1", the next third as group "2", and the last third group "3", as was done in Figure 7.

When the interface notices that the number of treatment groups does not evenly divide the number of data points, it reports this situation. Once the warning is acknowledged the interface will continue as best it can. Usually the user will want an equal number of points in each treatment group. If this is not the case, then using the plot by treatment group feature is not recommended as the interface may give unexpected results.

The input data file can be selected by choosing the menu items "File" then "Open DATA" (Figure 1). This opens the file viewer window so the user can select an input file. The interface opens the file viewer automatically if the user forgets to open the input file before starting computations.

2.2 Description Files

Optionally, a description file can be provided. This file can describe the individual features, giving them names, letting them be treated as discrete or continuous, and letting some of them be excluded from the analysis.

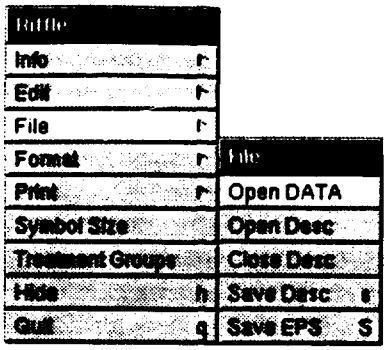


Figure 1: Main menu and File submenu.

Each feature name is required to be a string without blanks. As an example "Sepal_Length" with an underscore separating the words is acceptable, but "Sepal Length" separated by a space is not acceptable. The file should consist only of feature names, one of the words "`exclude`" or "`include`", and one of the words "`continuous`" or "`discrete`" on each line. Omitted descriptions default to "`include`" and "`continuous`". Here is a legal, but sloppy, example of such a file:

```
Sepal_Length   exclude      continuous
Sepal_Width    continuous
Petal_Length   continuous   include
Petal_Width    discrete     include
```

If no description file exists, the number of features is set equal to the number of data values in the first line of the data file, in which case all features are taken to be continuous and all are included.

The description file is opened by the file submenu item "Open Desc". Description files can also be closed or saved by items on the file submenu. Closing a description file will scan the data file to determine the number of features, reset the feature names to the defaults "Attr1", "Attr2", and so on, and will reset the features to include and continuous. This is appropriate when using a new data file that is not accurately represented by the last description file. Saving the description information to a file allows the user to retrieve that information when using the same data set or one with identical format.

2.3 Input Arguments

When performing a clustering computation the program requires information from the user. The Arguments window (Figure 2) accepts the user's choices, starts computations, graphs the data and results, and reports on the computation's statistical significance.

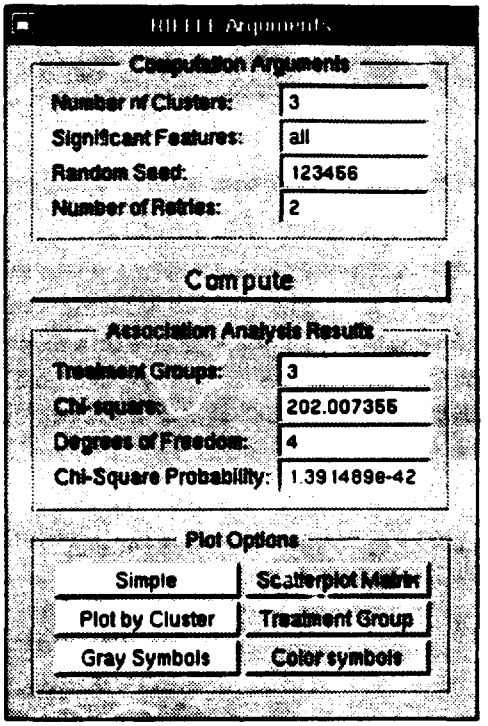


Figure 2: Arguments window.

2.3.1 Number of Clusters

The first argument is the number of clusters the program will fit the data to. A researcher may, or may not, know how many clusters are appropriate. Performing the computations with different numbers of clusters can give the user a feel for whether the data can be usefully described by clusters, and if so, how many clusters. The interface's text output (Section 2.5) will show

the average quality of each clustering run, giving the researcher an idea of how many clusters are most appropriate.

The researcher must, however, analyse the data further to verify the initial findings. For example, suppose a researcher starts with a value of three, then proceeds to four, five, and six clusters. The interface may show four clusters as having the highest quality, suggesting that there are four clusters in the data. However looking at the data may show that there are obviously only two clusters. How can this happen? Since four is the closest multiple of two, there is a good chance that four clusters will also show a strong quality, and may be misleading.

2.3.2 Significant Features

Significant features tells the program how many of the features to include in the computation. By stating "all" the program knows that every data feature is important. This input argument exploits the program's ability to automatically exclude some of the weaker features from consideration. For example, if there are six features for each data point, the researcher can ask the program to choose the best four features (Significant Features: [4]). This has the effect of excluding from the computation the two features that the algorithm finds contribute the least to the proportional reduction in error—the quality measure RIFFLE uses for clustering [5].

2.3.3 Random Seed

Each time a computation is run with the same information in the arguments window, and the same input file, the results will be identical. By changing the random seed the user can force the program to use a new set of pseudo-random numbers causing the results of the next computation to be different.

2.3.4 Number of Retries

The clustering job that RIFFLE is attempting to perform is an enormous task. Systematically checking every permutation of data points across the clusters would result in a computation that takes an intolerable length of time for all but the smallest data sets. In light of this, clustering algorithms, including RIFFLE, make approximations to this ideal. RIFFLE uses pseudo-random numbers to place the points in initial clusters, then proceeds to rearrange the points until a local best clustering case is arrived at. The

"Number of Retries" is the number of times RIFFLE is to perform this analysis, each time keeping the results only if the overall quality was better than the previous best.

So what values are appropriate inputs? Ten retries usually gives excellent results, and good results can be obtained in five or fewer retries for a quick analysis. If, for instance, you wish to simply look at the data plots with less emphasis on how well the points are clustered, then one retry is all you need.

2.4 Computing

The Compute button will begin the computation. If the input file has already been selected the computation will start right away. If the input has not been opened, the interface will open a file viewer window that allows the user to select the input data file. When the computation is done the output will be displayed in text form in the Results window (Section 2.5), the association analysis values (Section 2.6) will be displayed, the Features window will be updated (Section 2.8), and the results will be plotted (Section 2.7), .

2.5 Text Results

The results of the computation are displayed in the Results window as shown in Figure 3. The text results include the file name, and most of the data and computation results in text form so that these results can be printed, capturing the important aspects.

In this example the data file "iris.dat" was analyzed. The next line indicates that 150 data points with four features ("attributes") were clustered into three groups, using all four features.

The text results also report the number of features that the program found suspicious of having degenerate data. A degenerate feature might have an excessive percentage of identical data values. When these instances are found, the program does not use these features in the computation, and marks them as excluded in the features window.

Next, the χ^2 (chi-square) statistics are listed, as discussed in Section 2.6.

"Qual" is the quality, or proportional reduction in error (PRE) for that attribute, as discussed in Matthews and Hearne [5].

The ranks and values are the ranks (in a list of the values of that feature, **sorted in descending order**), and actual values of the data points used for that split. For example, in Figure 3 the Sepal_Length line indicates that

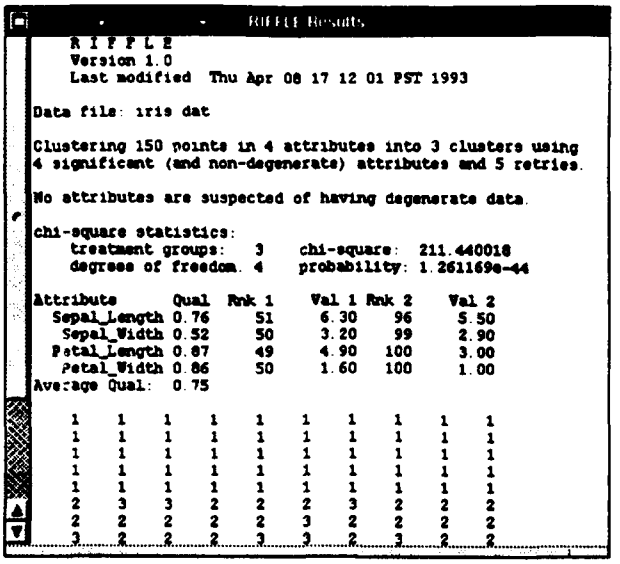


Figure 3: Results window showing text output.

the first split for that feature is at the data point with the 51st largest sepal length (out of the 150 member sample), and the actual data value for that 51st largest sepal length is 6.30. Likewise the same line shows the second split is at the 96th largest sepal length (out of 150), and the actual value is 5.50.

If we are looking for two clusters then one split point is defined, and the results show one rank "Rnk" column and one value "Val" column. If we are looking for three clusters then two split points are defined, each split point having one set of rank and value columns.

The numbers at the end of the text output are the cluster numbers for each point. Figure 3 shows only some of the 150 cluster numbers, the rest can be viewed by scrolling the lower part of the text field into view. The top left cluster number is the first data point in the file, the next one to the right is the second data point, and so on.

The cluster numbering is arbitrary and will change from run to run even though the overall pattern of the clusters will typically remain the same. For instance, with a single retry the first data point may be in cluster "3". Rerunning the program with two retries may label that same data point

(and for that matter, most of the other points in it's cluster) with number "2". The results are probably very similar, except that the cluster that was labeled "3" the first time is now labeled as cluster "2". This also happens with the labels (graph symbols) in the plots.

If the program is run with fewer significant features than total features then a value of 0.0 will be shown in the quality column of the non-significant features.

If some features are excluded by the description file, or by the features window, those features will not show up on the text output at all.

Menu items in the Interface are available to print the text results to paper or save them to a file.

2.6 Association Analysis Results

The association analysis statistics appear in the arguments window. Figure 4 shows the results for the analysis on the iris data.

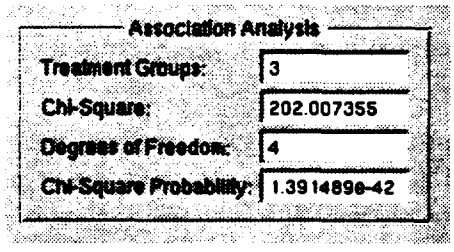


Figure 4: Association analysis of the iris data.

The plot by treatment group section (2.7.4), gives a discussion on the use of treatment groups. The number of treatment groups field is set equal to the number of clusters selected. However, plotting by treatment group, which requires that the number of treatment groups evenly divide the number of data points, may not give meaningful results for all values of clusters (and treatment groups).

In order for the association analysis, and plot by treatment group features to work the data file must be structured in a way that allows the interface to distinguish the groups. See Section 2.1 for information on the correct data file structure.

The interface uses the χ^2 (chi-square) statistic to show the significance of the association between the two groupings (*treatment groups* and *clusters*). The null hypothesis is that *treatment groups* and *cluster numbers* have no association. In this case the probability of a particular value of cluster number given a particular value of treatment group should be the same as the probability of that value of cluster number regardless of treatment group. Small values of probability indicate a significant association [9]. The probability value is the probability that there is *no association* (Since the null hypothesis states that there is no association). Values that are below 0.01, for example, indicate that there is a greater than 99% probability that a significant association exists between the clusters and treatment groups.

Notice that the probability values are frequently shown in scientific notation (i.e. 1.391489e–42) and are always between zero and one.

2.7 Graph Results

The interface also displays the clustering results graphically. This is a way of representing, at the same time, both the input data and the cluster assignments listed at the bottom of the results window.

The interface can graphically display the data by simple plot, or scatterplot matrix, and by cluster, or treatment group.

All of the graphs have the property that more than one point can be at the same location, causing the symbols to be plotted one on top of the other. This may result in the underneath point being invisible, and may result in plots that have fewer data points than expected. This may also result in unusual symbols. For example Figure 5 shows a few square symbols which are filled in with black, and yet there are only three clusters each with a single symbol: a black disk, a gray disk, and an empty black square frame. The fourth symbol is a black disk with a square plotted at the same location. If however, the symbols are exactly the same, or even just the same shape, then two points at the same location will show only the symbol plotted last.

The same is true for plots that use numerals and letters to represent points. If different numerals are plotted at the same location, they will result in an unusual symbol. But if two occurrences of the same numeral occupy the same location they will be indistinguishable from one point occupying that location.

As with the text results, menu items in the interface can print the plots to paper, or save the plots to an encapsulated PostScript file (.eps).

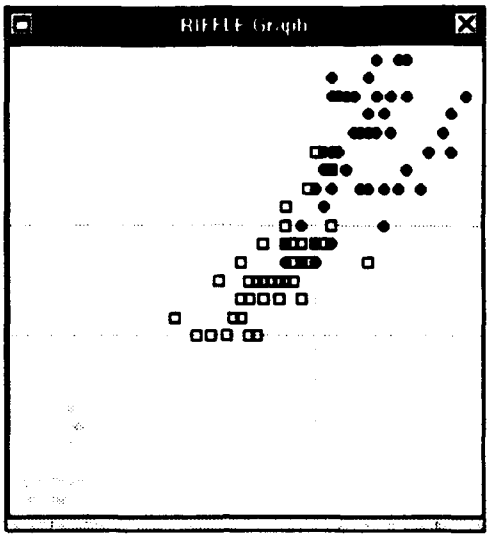


Figure 5: Simple Plot of Anderson's iris data (two best features showing three clusters).

In order for the data points to be plotted they must be "included" in the Features window (see Section 2.8).

2.7.1 Simple Plot

The simple plot shows any two **included** features graphed against each other in two dimensions (Figure 5). The x and y features are selected in the features window (Section 2.8). At the end of the computation the interface figures out which features are the two best (those with the highest PRE values) and automatically sets the best feature to the x axis and the second best to the y axis.

The vertical and horizontal lines show the split points for the clustering. Figure 5 shows two split points on each axis because three clusters were requested. At least one point will always fall on each split. When a point lies on the split it indicates that the point belongs to the region above (if the line is horizontal), or to the right (if the line is vertical). In Figure 5 the squares on the bottom and left edges of their cluster are shown on the split lines, but are in fact included in the center cell of the nine cell grid.

The simple plot can give the user an intuitive feel for how well the data fit into clusters. Figure 5 suggests a fairly strong clustering since both x and y features predict almost perfectly which group a point is in.

It should be noted that the interface can only plot a limited number of cluster symbols. At this writing the limit is fifteen, however, the symbols become distracting when there are too many. The plots have a better appearance with a fewer number of cluster symbols. Also, at some point the symbols are exhausted and letters are used as symbols. Similarly, the number of treatments groups is currently limited to fifteen. If more than fifteen clusters are requested, the data points placed in clusters sixteen and higher will all be plotted with the symbol "?".

2.7.2 Scatterplot Matrix

The scatterplot matrix shows several features plotted against each other in two dimensions (Figure 6). At the current time up to six features can be included in the matrix.

As it does for the simple plot, the interface finds the best set of features automatically (based on PRE values) for display in the scatterplot matrix. Other features can be selected with the Features window and the matrix will automatically resize to accommodate fewer or more features (up to the maximum), without changing the window size.

The S.W. to N.E. diagonal is filled with text cells. Each text cell indicates that plots on the same **row** use that feature on the **y** axis, and plots on the same **column** use that feature on the **x** axis. Centered vertically and horizontally in the text cell are the feature name and quality. In the S.W. corner of each text cell is the input file's minimum value for that feature, and in the N.E. corner is the maximum value.

The example in Figure 6 shows that two of the features are predictive (with a quality value close to the maximum of 1.0), petal length and petal width. It also shows that the sepal width feature does not contribute much to this particular clustering since the data points do not separate into discernible groups along that axis in the matrix. This visual weakness reaffirms the feature's low quality value.

Observe that the scatterplot matrix's top row, third column is the same plot as that shown in Figure 5, except that it is scaled differently (fitting the matrix into the given window dimensions).

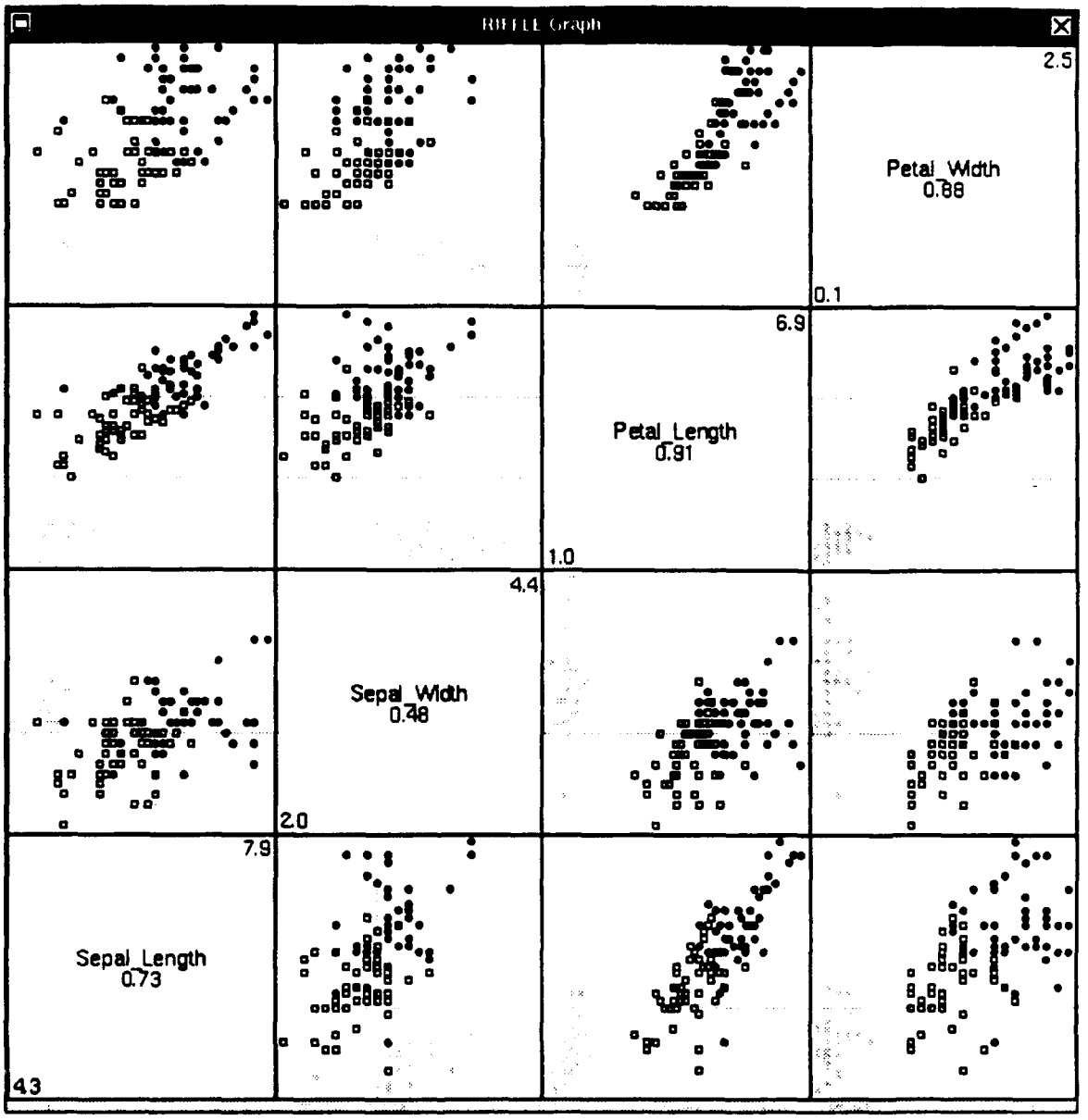


Figure 6: Scatterplot Matrix showing all four Iris features.

2.7.3 Plot by Cluster

Plotting by Cluster is demonstrated in the two sections above. In these cases the plots show how the RIFFLE program places the data points into clusters. Plotting by cluster represents each cluster by a different geometric symbol or color, whereas plotting by treatment group represents each group by numeral. The plot by cluster option will be understood better by examining its converse, plot by treatment group.

2.7.4 Plot by Treatment Group

Figure 7 shows the same data as the simple plot of Figure 5 with the plot by treatment group option instead of the plot by cluster option. Figure 7 is scaled to have a longer x axis than Figure 5 to more clearly separate the points. The points in plot by treatment group are indicated by numerals instead of the geometric symbols and colors used in plotting by cluster.

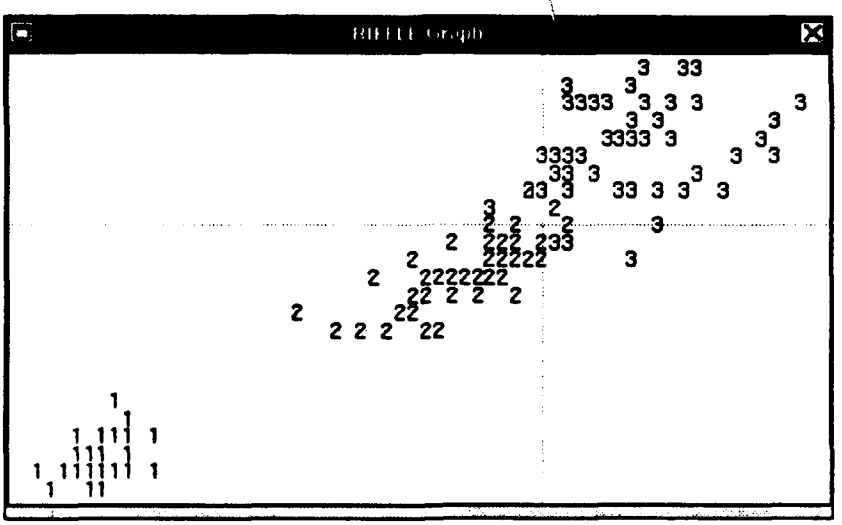


Figure 7: Simple Plot by treatment group

Other than the labels, the graphing is done the same way the plot by cluster is done so a direct comparison is appropriate. In fact, it is anticipated that researchers will swap between the two plot types to check for differences between the treatment groups and clusters.

Graphing with this option is a tool to help answer the question, "Now that I have the results of the clustering, how closely do they match the subpopulations that I know exist?" In the Iris example there would be three treatment groups which correspond to the three types of irises studied. The question is, "Do the clusters do a good job of grouping the data by the type of iris?" Swapping between the two plot types, with the same scale, can help to answer this question.

The RIFFLE program is "blind" to the treatment groups. That is, RIFFLE assigns points to clusters without any knowledge of which treatment group the points come from. However, the data is plotted by the interface, not RIFFLE. The interface uses structure in the input file to show treatment groups while RIFFLE remains naïve to that information.

In order to plot by treatment group, the data file must be structured in a way that allows the interface to distinguish the groups. As discussed in Section 2.1,

- all of the points in a group must be listed consecutively in the data file, and
- the groups must have an equal number of points.

The user indicates the number of groups in the data by using the arguments window. With three groups, the interface will number the first third of the data points "1", the next third "2", and the last third "3", as was done in Figure 7.

If the number of treatment groups does not evenly divide the number of data points, the interface will report this situation. Once the warning is acknowledged the interface will render the plot. Usually the user will want an equal number of points in each treatment group. If this is not the case, the interface may give unexpected results.

2.7.5 Adjustable Symbol Size

Symbol size in the plots can be adjusted with the menu item of the same name. At this time the plot by treatment group option does not support the adjustable numeral size, but Plot by Cluster can adjust its symbol size. The best size will typically depend on the number of points in the dataset, and the size of the plots (which can be changed by resizing the window). Using one size for the simple plot and a slightly smaller size for the scatterplot matrix seems to work well.

2.7.6 Printing Graphs

The Print menu item has both "Graph", and "Full Page Graph" options. The Graph option will print the graph window at it's current size. If the window is larger than one page then the printing process gives unpredictable results. Otherwise the printed graph will be approximately the size seen on the screen. The Full Page Graph option resizes the viewing window to page size, leaving half inch margins, and then directs the output to the printer. This option will make printed graphs with the graph scaled to page size, and would be useful for making printed graphs that are always the same dimensions (i.e. not dependant on how you resized the graph window in that particular session). It is best to choose portrait or landscape page orientation with the Format menu item prior to using the Full Page Graph option.

2.8 Features Window

The Features window provides on the fly choices paralleling those made with the description file, and also controls which features are graphed in the simple plot and scatterplot matrix.

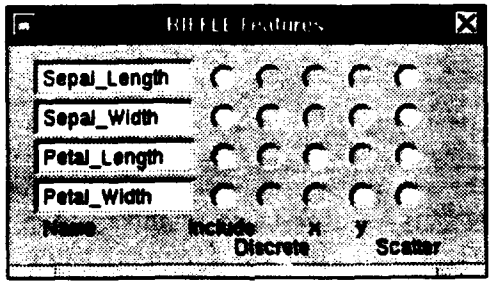


Figure 8: Features window: highlighted buttons indicate features included in computation, and features to graph in simple (x and y) and matrix (Scatter) plots.

The features window allows the user to change the feature name, include or exclude the feature in the computation, and designate the feature as either continuous or discrete. The features window also controls which features are plotted in the simple and scatterplot graphs. Columns "x" and "y" allow exactly one of the features to be selected at any time. These columns direct

which features are plotted on the simple plot's x, and y axes. The Scatter column, on the other hand, allows two or more features to be selected, and will plot these in the scatterplot matrix (up to the maximum).

Figure 8 shows all four features included in the computation, all features are continuous, the third and fourth features are selected for the simple plot (columns x and y), and all features are selected for the scatterplot matrix (column Scatter).

A feature must be **included** in the computation in order for the interface to plot it. The interface will edit for this requirement and un-select the feature if it does not qualify for plotting. This will cause an error panel to appear, and the graph will be cleared. It should be noted that the interface will crash if the feature is excluded, computed, included, and then graphed.

The information in the Features window (feature name, whether it is included or excluded, and whether it is continuous or discrete) can be saved to a description file (Section 2.2) by the menu "File", submenu "Save Desc".

2.9 Color

The color button allows the graphs to display different clusters by color, always using the same symbol shape. Even if color is used on the screen, when printing, the interface will adjust the symbols to accommodate a non-color printer.

3 Future Plans

Although clustering has historically been considered an exploratory data analysis technique, the research team is investigating promising applications of the nonmetric clustering tool for predictive statistics as well. The team is developing an interface version that includes tools for performing a broader cross-analysis of treatment group type data with several statistical techniques including the RIFFLE algorithm.

4 More Information

For more information about nonmetric clustering, the RIFFLE program, or their applications refer to these papers [5, 6, 7, 8, 9]. Questions about the interface, or the above issues can also be directed to:

Michael J. Roze:	`roze@rum.cs.wwu.edu`
Geoffrey B. Matthews:	`matthews@fortress.cs.wwu.edu`.

5 Acknowledgement

This research is supported by USAFOSR Grant No. AFOSR-91-2091 DEF.

References

[1] E. Anderson. The irises of the Gaspe Peninsula. *Bull. Amer. Iris Soc.*, 59:2–5, 1935.

[2] E. Anderson. The species problem in Iris. *Ann. Mo. bot. Gdn.*, 3:511–25, 1936.

[3] E. Anderson. A semigraphical method for the analysis of complex problems. *Proc. of the Nat. Academy of Sciences*, 13:923–27, 1957.

[4] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, 7:179–88, 1936. Reprinted in Fisher, R. A. (1950). em Contributions to Mathematical Statistics. Wiley, New York.

[5] Geoffrey Matthews and James Hearne. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175–184, 1991.

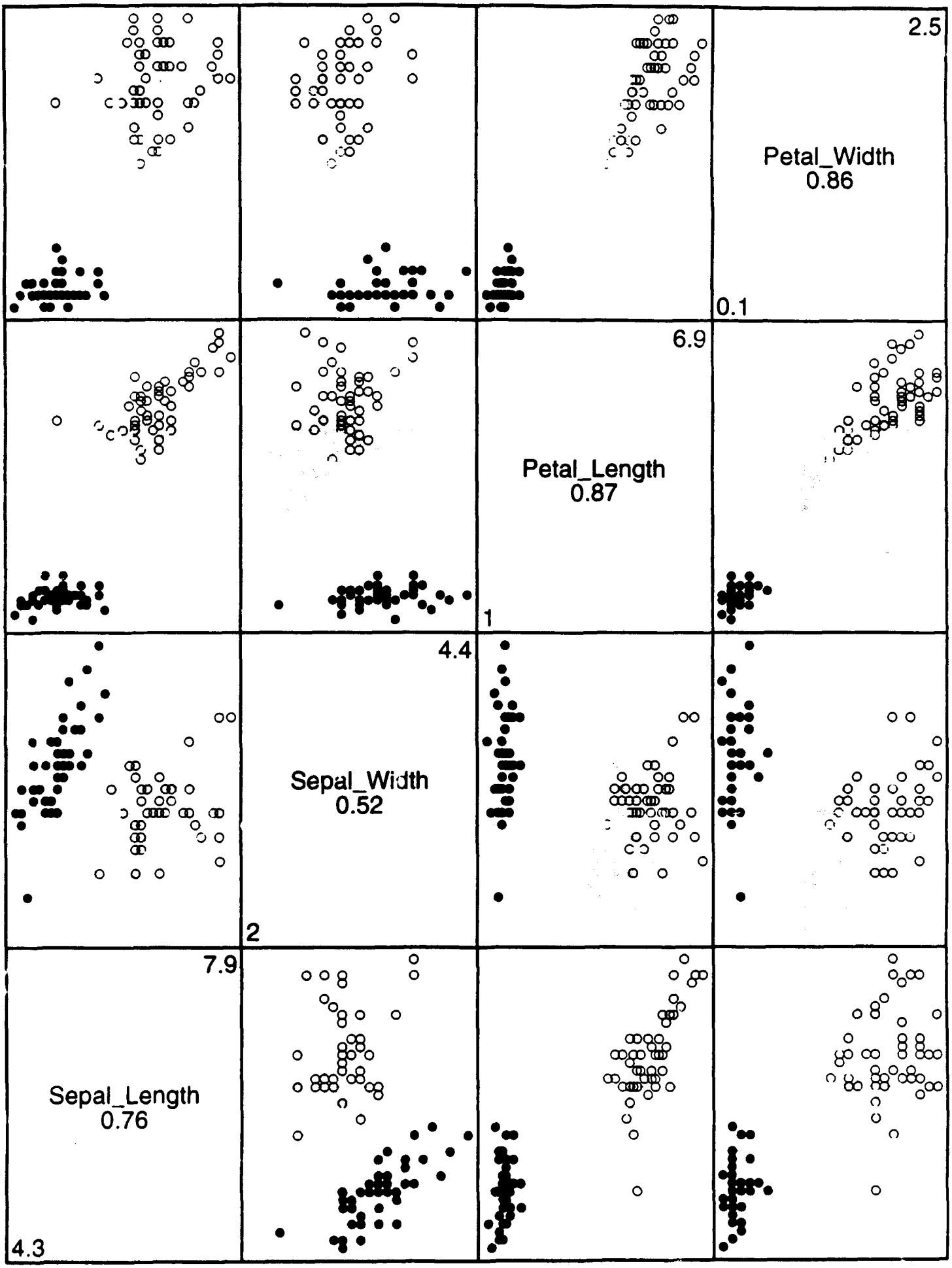
[6] Geoffrey Matthews and Robin Matthews. A model for describing community change. In *Pesticides in Natural Systems: How Can Their Effects Be Monitored?*, Environmental Research Laboratory/CRD, Corvallis Oregon, 1990.

[7] Geoffrey B. Matthews, Robin A. Matthews, and Barnard Hachmöller. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences*, forthcoming, 1991.

[8] Robin A. Matthews, Geoffrey B. Matthews, and William J. Ehinger. Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modelling*, 53, 1991.

[9] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1988.

[10] Eric P. Smith, Kurt W. Pontasch, and John Cairns, Jr. Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.*, 24(4):507–514, 1990.


2.5
Petal_Width
0.86
0.1
6.9
Petal_Length
0.87
1
4.4
Sepal_Width
0.52
2
7.9
Sepal_Length
0.76
4.3

Environmental Sciences, 2, 3 (1993) 113-130
MYU, Tokyo

ES031

Multivariate Analyses of the Impact of the Turbine Fuel Jet-A Using a Standard Aquatic Microcosm Toxicity Test

W. G. Landis, R. A. Matthews, A. J. Markiewicz, N. J. Shough and G. B. Matthews*

Institute of Environmental Toxicology and Chemistry
Huxley College of Environmental Studies
Western Washington University
Bellingham, WA 98225 USA
*Computer Science Department
Western Washington University
Bellingham, WA 98225 USA

(Received September 9, 1992; accepted July 24, 1993)

Key words: multispecies toxicity testing, Jet-A, microcosm, nonmetric clustering, multivariate analysis

We investigated the toxicity of the water soluble fraction (WSF) of the turbine fuel Jet-A using the standard aquatic microcosm (SAM) method. The SAM experiment was conducted using concentrations of 0, 1, 5 and 15% WSF in 3 L SAMs containing 14 species of organisms. The toxicant was added on day 7 of the 63-day experiment. Physical, chemical, and biological measurements were collected twice each week from day 11 through day 63. In the highest WSF treatment group an algal bloom ensued, generated by the toxicity of the WSF to *Daphnia*. As the test proceeded, the *Daphnia* populations increased and the algal populations decreased to about the reference values. In the last few weeks of the experiment *Cyprinotus* (ostracod) densities were higher in the reference than in the other treatment groups and *Philodina* (rotifer) densities were lower in the reference than in the other treatment groups. Because of high sampling variance, the ANOVA results suggested that few of these effects were significant. Multivariate analyses, however, revealed two distinct divergences between treatment groups: an early divergence that was probably due to the *Daphnia*/algae response, and a late divergence that was much more subtle, and may have been related to changes in the detrital quality in the different treatment groups. The variables that were most important in distinguishing the four treatments shifted during the course of the experiment, demonstrating the fallacy of

using only one index or a few measured endpoints in the evaluation of community-level interactions.

1. Introduction

Multispecies toxicity tests are usually referred to as microcosm or mesocosm tests, although a clear definition of these terms has not been put forth. Multispecies toxicity test systems range from approximately 1 L (e.g., mixed flask cultures) to thousands of liters, as in the case of the pond mesocosms used in pesticide registration testing. In the standardized aquatic microcosm (SAM) method[1] developed by Taub and colleagues,[2-12] the composition of the microcosm is clearly defined (Table 1). In other types of microcosms, the physical, chemical, and biological composi-

Table 1
Summary of test conditions for conducting the SAM Jet-A toxicity test.

Organisms:	Algae added on day 0 at 10^3 cells for each taxon: *Anabaena cylindrica*, *Ankistrodesmus* sp., *Chlamydomonas reinhardi* 90, *Chlorella vulgaris*, *Lyngbya* sp., *Scenedesmus obliquus*, *Selenastrum capricornutum*, *Stigeoclonium* sp., and *Ulothrix* sp. Animals added on day 4 at concentrations in parentheses: *Daphnia magna* (16), *Cypridopsis* sp (ostracod) (6), *Hypotricha* (protozoa) (0.1/ml), *Philodina* sp. (rotifer) (0.03/ml)
Test vessel:	One-gallon (3.8 L) glass jars; 16.0 cm wide at the shoulder; 25 cm tall with 10.6 cm openings
Medium:	T82MV; 3 L added to each container
Sediment:	Autoclaved silica sand (200 g), ground, crude chitin (0.5 g), and cellulose powder (0.5 g) added to each container
Replication:	6 replicate microcosms × 4 treatments
Reinoculation: (each microcosm)	Once per week one drop (~0.05 ml) added to each microcosm from a mix containing 5×10^2 cells of each alga
Addition of test materials:	Test material added on day 7 by removing 450 ml from each container and then adding appropriate amounts of the WSF to produce concentrations of 0, 1, 5 and 15 percent WSF. After toxicant addition the final volume was adjusted to 3 L
Test duration:	63 days
Temperature:	20° to 25°C
Light intensity:	80 μE m^2 photosynthetically active radiation/s (850 to 1000 fc)
Photoperiod:	12 h light/12 h dark
Sampling frequency:	2 times each week
Measurements:	Algal, invertebrate and protozoa counts, pH, dissolved oxygen, optical density. Calculated parameters included species concentrations, DO, DO gain and loss, net P/R ratio, pH, algal species diversity, *Daphnia* fecundity, algal biovolume, and biovolume of available algae

tions may vary widely.

Typically, the goals of multispecies toxicity tests are to detect changes in the population dynamics of the individual taxa that would not be apparent in single-species tests, and to detect community-level differences that are correlated with treatment groups. One of the major difficulties in the evaluation of multispecies toxicity tests has been to analyze the complex data set on a level consistent with these goals. A number of statistical approaches have been used to evaluate multispecies toxicity data. Analysis of variance (ANOVA) is the classic method used to examine differences between the treatment groups. However, because multispecies toxicity tests generally run for weeks, or even months, there are problems with using ANOVA, including the increased likelihood of a Type II error (accepting a false null-hypothesis), the presence of temporal dependence among the variables, and the difficulty of graphically representing the results. Conquest and Taub[13] developed a method to overcome some of the problems by using intervals of nonsignificant difference (INDs). This method corrects for the likelihood of Type II errors and produces intervals that are easily graphed. The method is routinely used to examine data from SAM toxicity tests, and is applicable to other multivariate toxicity tests. The major drawback is that this method can only be used to examine one variable at a time. While this addresses the first goal in multispecies toxicity testing, it ignores the second.

Multivariate data analysis methods are necessary to address the second goal of detecting community-level differences. One of the first multivariate methods used in toxicity testing was the calculation of ecosystem strain developed by Kersting[14–16] for a relatively simple (three species) microcosm. At about the same time, Johnson[17,18] developed a multivariate algorithm using the n-dimensional coordinates of a multivariate data set and the distances between these coordinates as a measure of divergence between treatment groups. Both of these methods have the advantage of examining the ecosystem as a whole rather than by single variables. A major disadvantage of both these multivariate methods (and of many others) is that all of the data are usually incorporated without regard to measurement units or the appropriateness of including all variables, even random ones, in the analysis.

Ideally, a multivariate statistical test used for evaluating complex data sets will have the following characteristics: (i) it will not combine counts from dissimilar taxa by means of sums of squares, or other *ad hoc* mathematical techniques; (ii) it will not require transformations of the data; (iii) it will work without modification on incomplete data sets; (iv) it will work without further assumptions on different data types (e.g., species counts or presence/absence data); (v) the significance of a taxon to the analysis will not depend on its abundance, so rare taxa can compete in importance with common taxa; (vi) it will provide an integral measure of "how good" the analysis is (i.e., whether the data set differs from a random collection of points); (vii) it will, in some cases, identify a subset of the taxa that serve as reliable indicators of the physical environment. To our knowledge, only one multivariate technique (nonmetric clustering) satisfies all these criteria.[19]

In this paper, we use ANOVA (with INDs) and three multivariate techniques to search for meaningful patterns in data from a SAM toxicity test using the water soluble fraction (WSF) of Jet-A turbine fuel. Jet-A is one of the most widely available aviation fuels, and, because of its stringent manufacturing specifications, is an excellent choice for evaluating the effects of a complex organic toxicant on a multispecies system. The multivariate techniques include two conventional tests based on the ratio of multivariate metric distances (Euclidean and cosine of the vector distances), and one relatively new procedure, nonmetric clustering and association analysis.[19] All three of the multivariate techniques have proven useful in analyzing complex ecological data sets.[20-22]

2. Materials and Methods

2.1 *Reagents*

All chemicals used in the culture of the organisms and in the formulation of the microcosm media were reagent grade or as specified in the ASTM protocol.[1] Glassware for the preparation of the WSF of Jet-A was washed in nonphosphate soap, rinsed, soaked in 2N HCl for at least 1 h, rinsed ten times with distilled water, dried, and autoclaved for 30 min. Jet-A was provided by Fliteline Services of Bellingham, Washington, U.S.A., and refined by Chevron. The sample was obtained from the sample valve used for quality control and water sampling to prevent contamination by the refueling apparatus. The shipment lot was recorded and is on file. Microcosm medium T82MV was used for extracting the soluble fraction of Jet-A. Twenty-five ml of Jet-A were added to a 1 L separatory funnel containing 1000 ml of T82MV medium. For 1 h, the mixture was repeatedly shaken for 5 min and allowed to stand for 15 min. The mixture was then allowed to stand overnight. The following day all but the upper 100 ml of the T82MV/WSF mixture was drained into a clean, sterile 1 L amber glass bottle and capped with a Teflon-lined screw cap. The WSF was used within 24 h or stored at 4°C for no longer than 48 h.

2.2 *Gas chromatography of WSF*

A gas chromatographic analysis of the WSF was carried out using a Tekmar LSC 2000 purge and trap (P&T) concentrator system in tandem with a Hewlett-Packard 5890A gas chromatograph and a flame ionization detector (FID).[23-25] Instrument blanks and deionized, distilled water blanks were used to verify the cleanliness of P&T and GC columns prior to analysis of the WSF samples. A 5 ml sample was injected into a 5 ml sparger, purged with prepurified nitrogen gas for 11 min and dry purged for 4 min. Volatile hydrocarbons, purged from the sample and collected on the Tenax/Silica gel column, were desorbed at 180°C directly onto the SPB-5 fused silica capillary column (30 m × 0.53 mm, ID 1.5 μm film). The column was held at 35°C for 2 min, increased to 225°C at 12°C/min, and held at that temperature for 5 min. A Spectra-Physics 4290 integrator was used to record the FID signal output of the volatile hydrocarbons that were separated and eluted from the

column by molecular weight.

2.3 *Short-term toxicity tests*

In order to determine the appropriate WSF concentrations to be used for the SAM microcosm, a series of short-term toxicity tests was performed. This included 96 h algal growth inhibition tests using three species of algae (*Chlamydamonas reinhardii, Ankistrodesmus falcatus*, and *Selenastrum capricornutum*) and a 48 h *Daphnia magna* acute toxicity test.

The test algae were grown in a semi-flow through culture apparatus on the microcosm media T82MV and collected during log-phase growth for inoculation into the test flasks. Five hundred ml Erlenmeyer flasks were used as test chambers. Each test chamber contained 100 ml of the following treatments (reps = 2/treatment): 0 (reference), 6.25, 12.5, 25, 50 and 100% WSF. All dilutions of the WSF were made using T82MV. The test organisms were added at a concentration of approximately 3.0×10^4 cells/ml. Test mixtures were incubated at 20.0°C ± 1.0°C, with a 12:12 h light/dark cycle. Cell densities were determined every 24 h during the 96 h test period using a Neubauer counting chamber. The cell numbers were plotted against the WSF concentrations. If possible, a least-squares regression line was drawn and the IC_{50} (concentration resulting in 50% inhibition compared to the control) was determined. Significant differences between groups were determined using ANOVA.

Daphnia magna 48 h acute toxicity tests[26] were conducted using T82MV medium at concentrations of 0, 6.25, 12.5, 25, 50 and 100% WSF (reps = 2/treatment). Ten neonates were placed in 250 ml beakers containing 100 ml of test solution. After 24 and 48 h, the numbers of dead cells were recorded. Data were analyzed graphically and statistically to obtain an estimate of the EC_{50}.

2.4 *SAM toxicity test*

The 63-day SAM protocol[1] was modified to allow dosing with the WSF. The WSF was added on day 7 by stirring each microcosm, removing 450 ml from each container, and adding WSF to produce concentrations of 0, 1, 5, and 15% WSF. The final volume was readjusted to 3 L using T82MV. An attempt was made to filter and retain the organisms withdrawn during the removal of the 450 ml prior to addition of the toxicant. All graphs and statistical analyses began with the next sampling day (day 11). Table 1 summarizes the organisms, conditions and modifications used for the Jet-A experiment.

2.5 *Data cnalysis*

The variables that were measured or calculated included the numerical densities for each species, dissolved oxygen (DO), DO gain and loss, net photosynthesis/respiration ratio (P/R), pH, algal species diversity, algal biovolume, and biovolume of "available" algae (i.e., available for consumption by filter feeders).[1] The ANOVA INDs[13] and the average values for each variable were plotted by treat-

ment group against time to identify significant differences. In addition, three multivariate clustering and significance tests were used to determine dose/response relationships. Two of the clustering procedures were based on the ratio of metric distances (Euclidean and cosine of vectors) within treatment groups vs between treatment groups. The third test used nonmetric clustering and association analysis.[19]

The biotic parameters used for the multivariate analyses are listed in Table 2. Treating each sameple on a given day as a vector of values, $x = \langle x_1 \cdots x_n \rangle$, with one value for each of the measured biotic variables, allows Euclidean distance between two sample points x and y to be computed as:

$$\sqrt{\sum_i (x_i - y_i)^2}.$$

The cosine of the vector distance between x and y can be computed as:

$$1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum y_i^2}}.$$

Subtracting the cosine from one yields a distance measure, rather than a similarity measure, with the measure increasing as the points get farther from each other.

The statistical significance of the metric clustering results was calculated using

Table 2

Biotic parameters used in the multivariate statistical tests.

Anabaena
Ankistrodesmus
Chlamydomonas
Chlorella
Daphnia
Ephipia
Small *Daphnia*
Medium *Daphnia*
Large *Daphnia*
Hypotricha (Protozoa)
Lyngbya
Miscellaneous sp.
Cyprinotus (Ostracod)
Philodina (Rotifer)
Scenedesmus
Selanastrum
Stigeoclonium
Ulothrix

Derived variables (e.g., diversity) were not used because they are not independent.

the within-between (W/B) ratio and an approximate randomization test.[29] For each date, one sample point x was obtained from each of six replicates in the four treatment groups, giving a 24 × 24 matrix of distances. After the distances were computed, the ratio of the average within group distance (W) to the average between group distance (B) was computed (W/B). If the points in a given treatment group were, on average, closer to each other than they were to points in a different treatment group, then this ratio will be small. The significance of the ratio was estimated using an approximate randomization test.[29] This test is based on the null hypothesis that assignment of points to treatment groups is random, the treatment having no effect. Accordingly, the test repeatedly (500 times) assigned the 24 points randomly to (pseudo) groups and calculated the W/B ratio. If the null hypothesis is false, the randomly derived W/B ratio will be larger, on average, than the W/B ratio obtained from the actual treatment groups. An estimate of the probability under the null hypothesis was obtained as $(n + 1)/(500 + 1)$, where n was the number of times the random W/B ratio was less than or equal to the actual W/B ratio.

In the nonmetric clustering and association test, the data were first clustered independently of treatment group, using the computer program RIFFLE.[22] Because the clustering analysis is naive to treatment group, the clusters may, or may not correspond to treatment effects. Under the null hypothesis, there should be no association between the clustering and the treatment groups. To test this hypothesis, the association between clusters and treatment groups was measured in a 4 × 4 contingency table, each point in treatment group i and cluster j being counted as a point in frequency cell ij. Significance of the association in the table was then measured with Pearson's χ^2 test:[30]

$$\chi^2 = \sum_{ij} \frac{(N_{ij} - n_{ij})^2}{n_{ij}}$$

$$n_{ij} = \frac{N_{+j} N_{i+}}{N},$$

where N_{ij} is the actual cell count; n_{ij} is the expected cell frequency obtained from the row (N_{+j}) and column (N_{i+}) marginal totals; and N is the total cell count (i.e., 24). The significance (probability under the null hypothesis) for this value of χ^2 was computed using standard procedures.[31]

3. Results

3.1 *GC analysis*

The results from the GC analysis of the WSF are shown in Fig. 1. Immediately after the WSF was added to the SAMS, approximately 50–60 peaks were distinguishable in the highest treatment group (15% WSF). By the end of the experiment, virtually all of the peaks had disappeared from the water column, probably

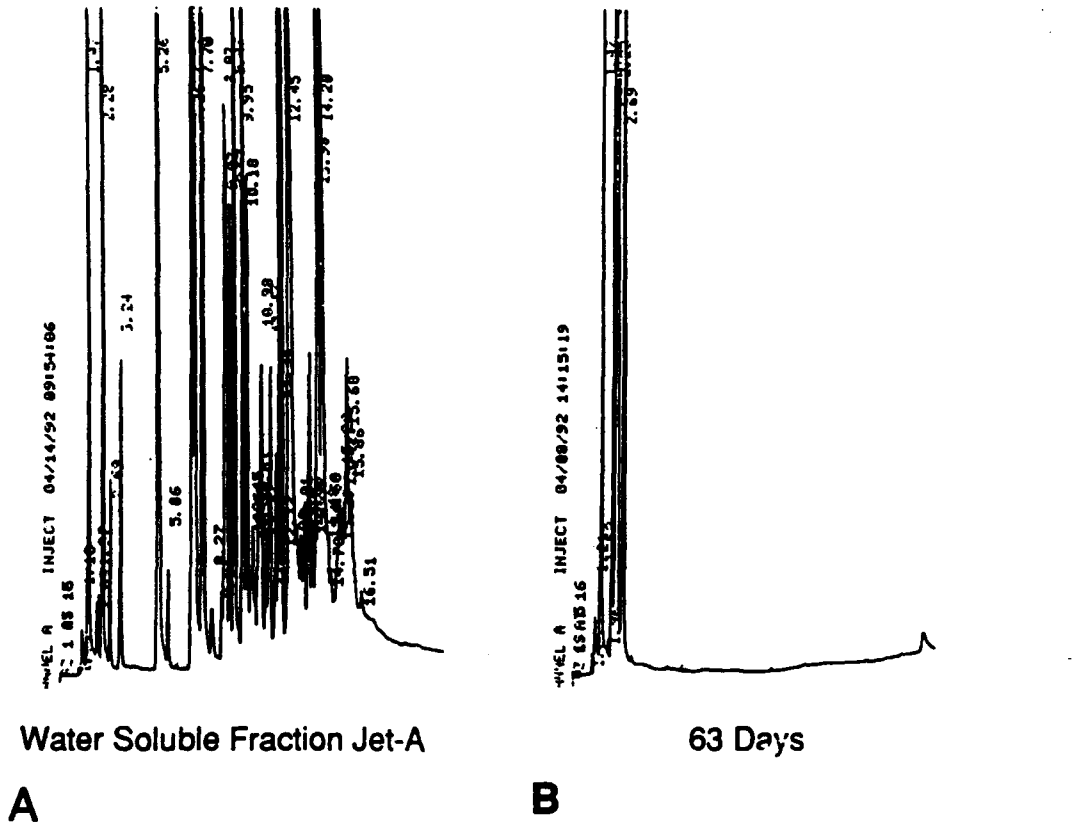


Fig. 1. Trap and purge GC chromatogram from the 15% WSF treatment group showing initial (day 11) and final (day 63) peaks.

due to volatization, photooxidation, biotransformation, and biodegradation.

3.2 *Short-term toxicity tests*

None of the 96 h acute algal toxicity tests indicated significant growth inhibition or enhancement correlated to treatment. However, the 48 h *D. magna* tests indicated that concentrations of 10–50% WSF caused *Daphnia* mortalities of 50–100%. The graphically derived EC_{50} was approximately 7% WSF (Fig. 2). Therefore, we expected that the highest concentration in the SAM experiments (15% WSF) would adversely impact the *Daphnia* populations shortly after the toxicant addition.

3.3 *SAM univariate results*

Daphnia population growth in the reference and lowest treatment group was similar throughout most of the experiment (Fig. 3). As expected, however, both of the higher treatment groups showed inhibition of *Daphnia* populations. In Treatment 3, the *Daphnia* populations (especially small *Daphnia*) started increasing on day 14. Treatment 4 did not show a major increase in the populations until day 17, and the population peak was not reached until after day 30.

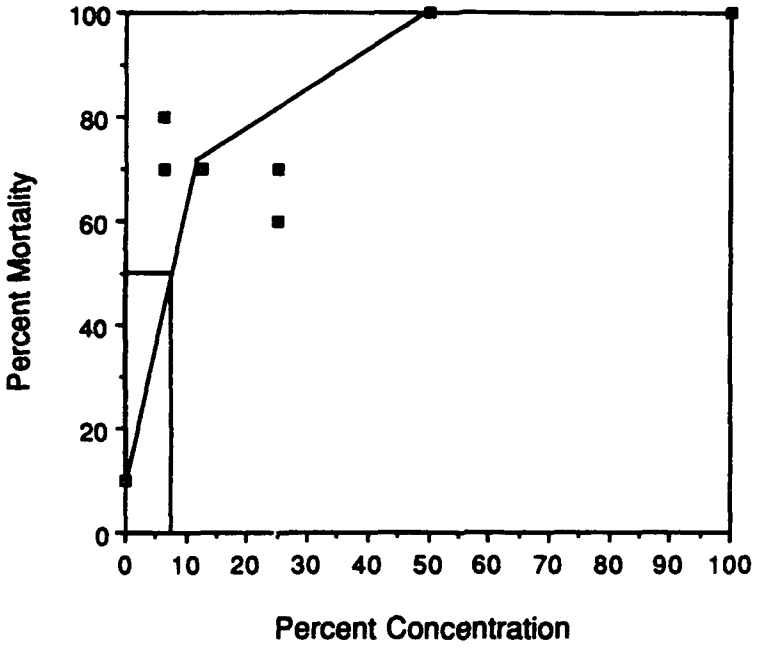


Fig. 2. Forty-eight h acute *Daphnia* toxicity tests results for the WSF of Jet-A.

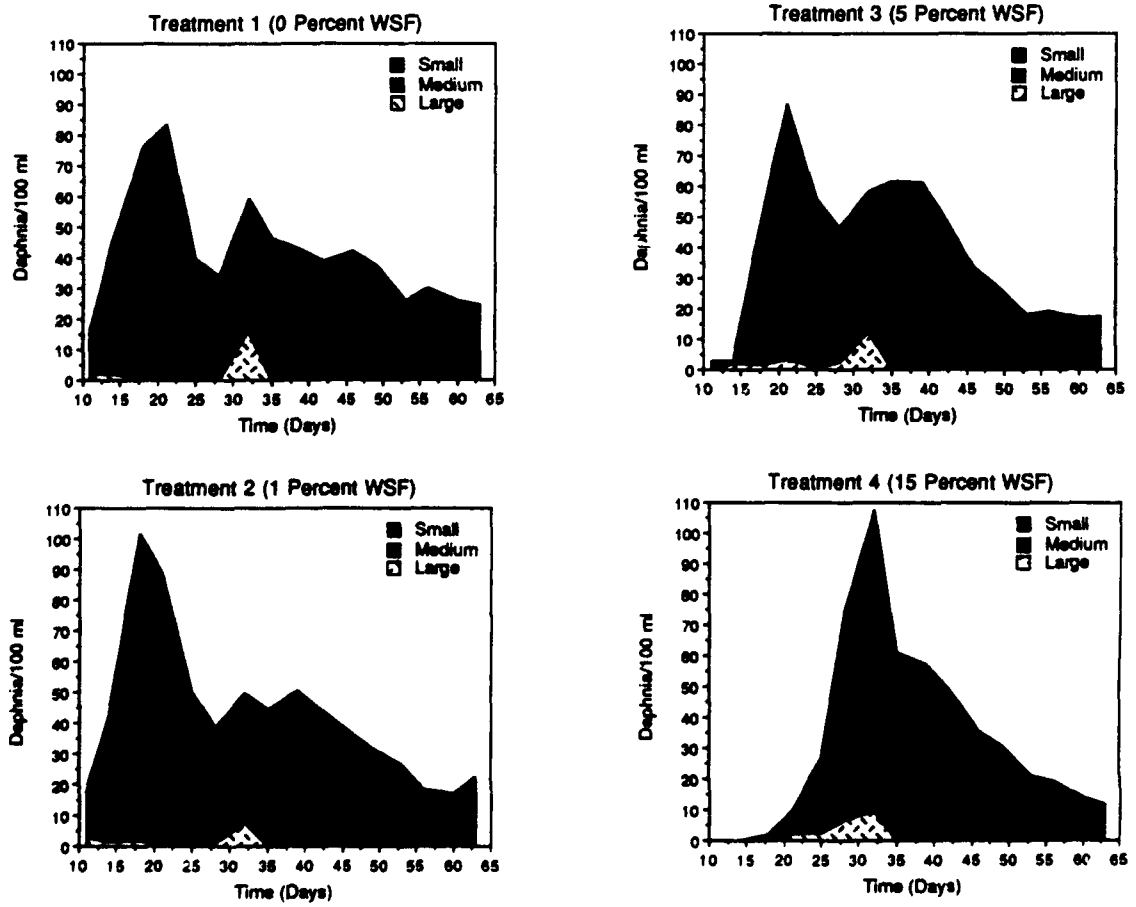


Fig. 3. *Daphnia magna* densities from the SAM toxicity test of the WSF of Jet-A.

Early algal blooms were observed in Treatments 3 and 4 (Fig. 4). On day 21 the peak algal density in Treatment 4 was approximately four times that of the reference. These increases were most likely due to reduced survival and reproduction in the *Daphnia* populations in the first few weeks of the experiment.

At the end of the experiment the average *Cyprinotus* (ostracod) density in the reference was approximately twice that of Treatment 4 (Fig. 5), and the population densities of other treatment groups were ranked in a dose/response manner. The ranking was consistent from day 49 onward. Because of the high sampling variance, the IND plots did not indicate any significant differences between treatments. Similarly, by the end of the experiment *Philodina* (rotifer), which were relatively uncommon throughout the experiment, were less numerous in the reference compared to Treatments 3 and 4. Again, because of the large sampling variance, the IND plots did not show any significant differences (Fig. 6).

The P/R ratio, measured by changes in daytime and nighttime DO concentrations, exhibited a dose/response relationship early in the experiment, with Treatments 3 and 4 being significantly different from the reference (Fig. 7A). The pH also responded in a dose/response manner to the addition of Jet-A. During the

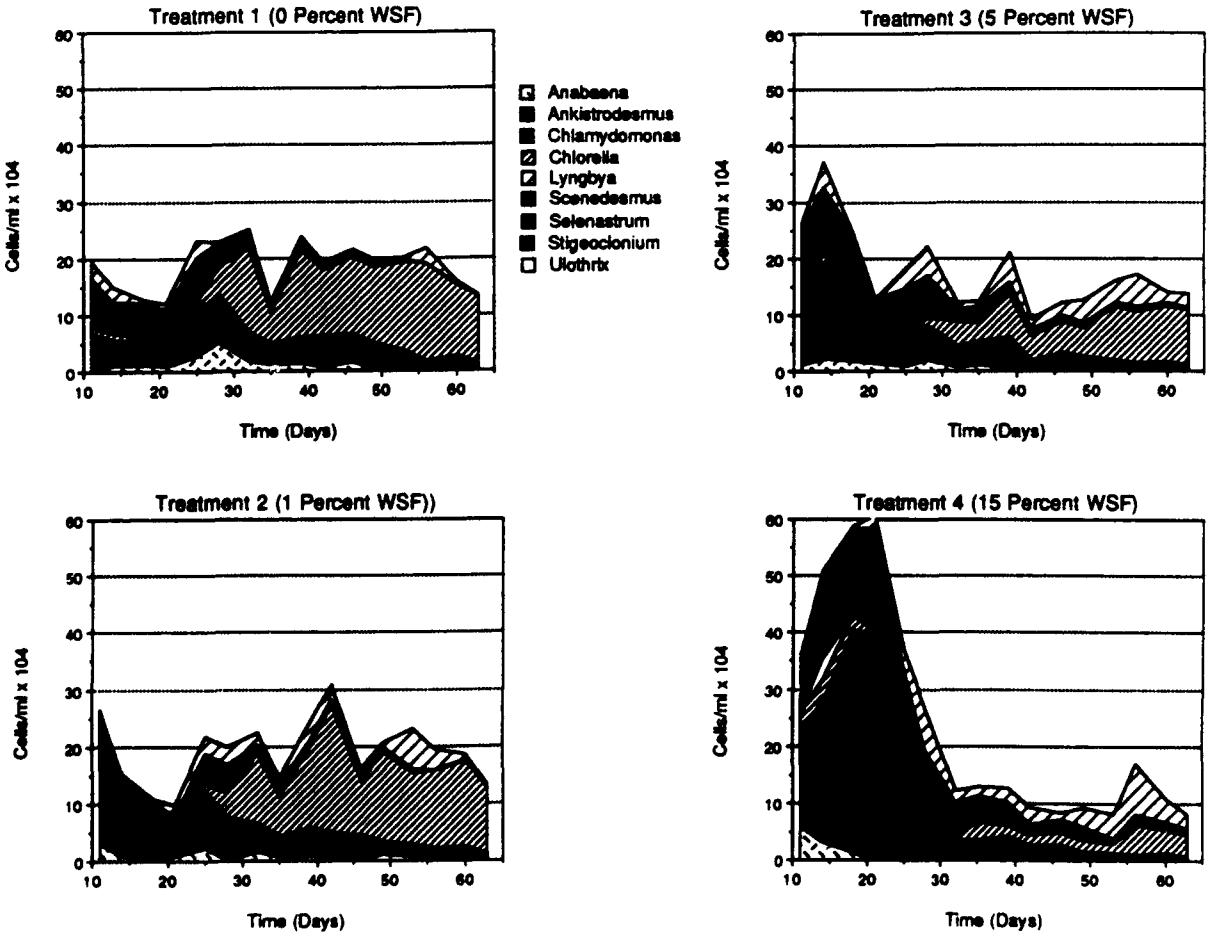


Fig. 4. Algal densities from the SAM toxicity test of the WSF of Jet-A.

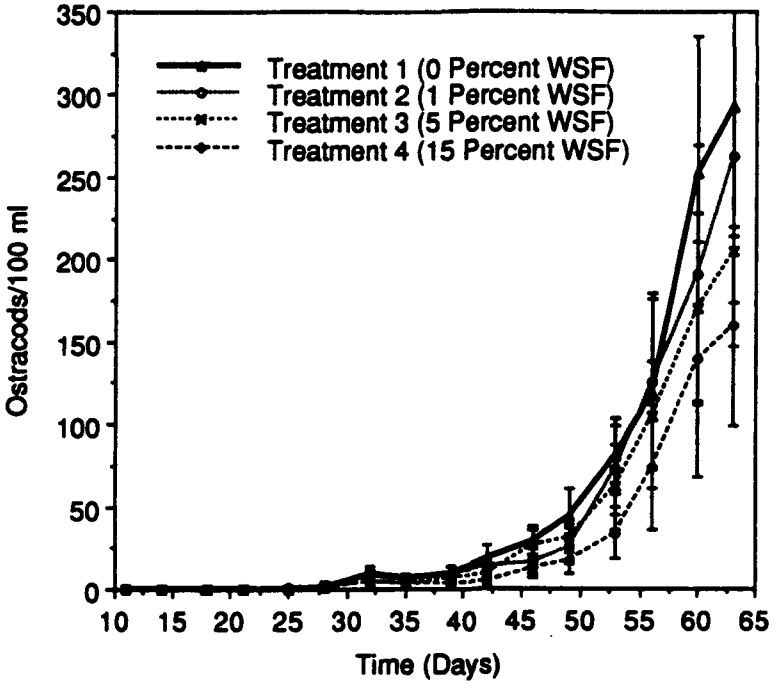


Fig. 5. *Cyprinotus* densities from the SAM toxicity test of the WSF of Jet-A.

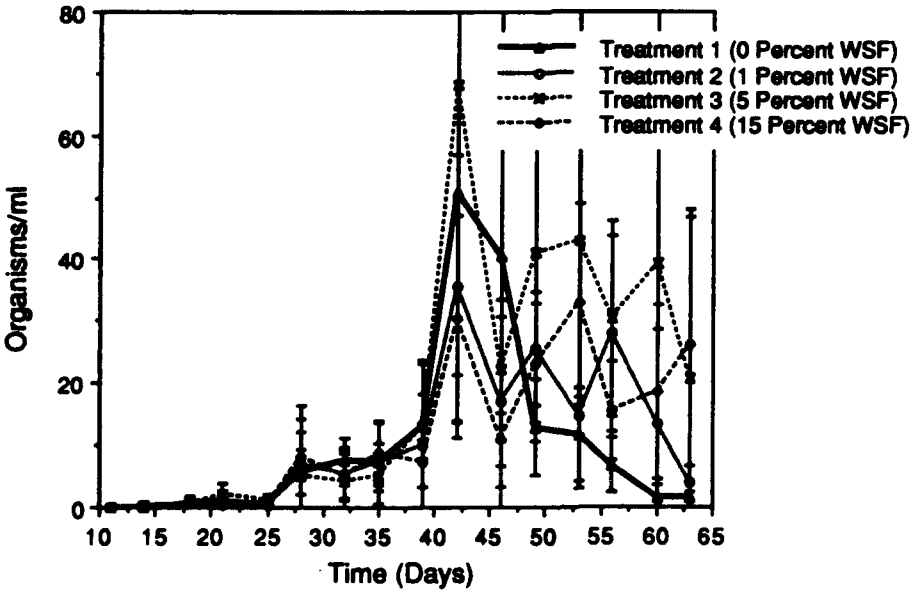


Fig. 6. *Philodina* densities from the SAM toxicity test of the WSF of Jet-A.

early part of the experiment (during the algal blooms), pH was significantly higher in the two highest treatment groups than in the reference (Fig. 7B). On day 49 a second deviation from the reference was detected. No significant differences in pH were observed among the treatment groups by the end of the experiment.

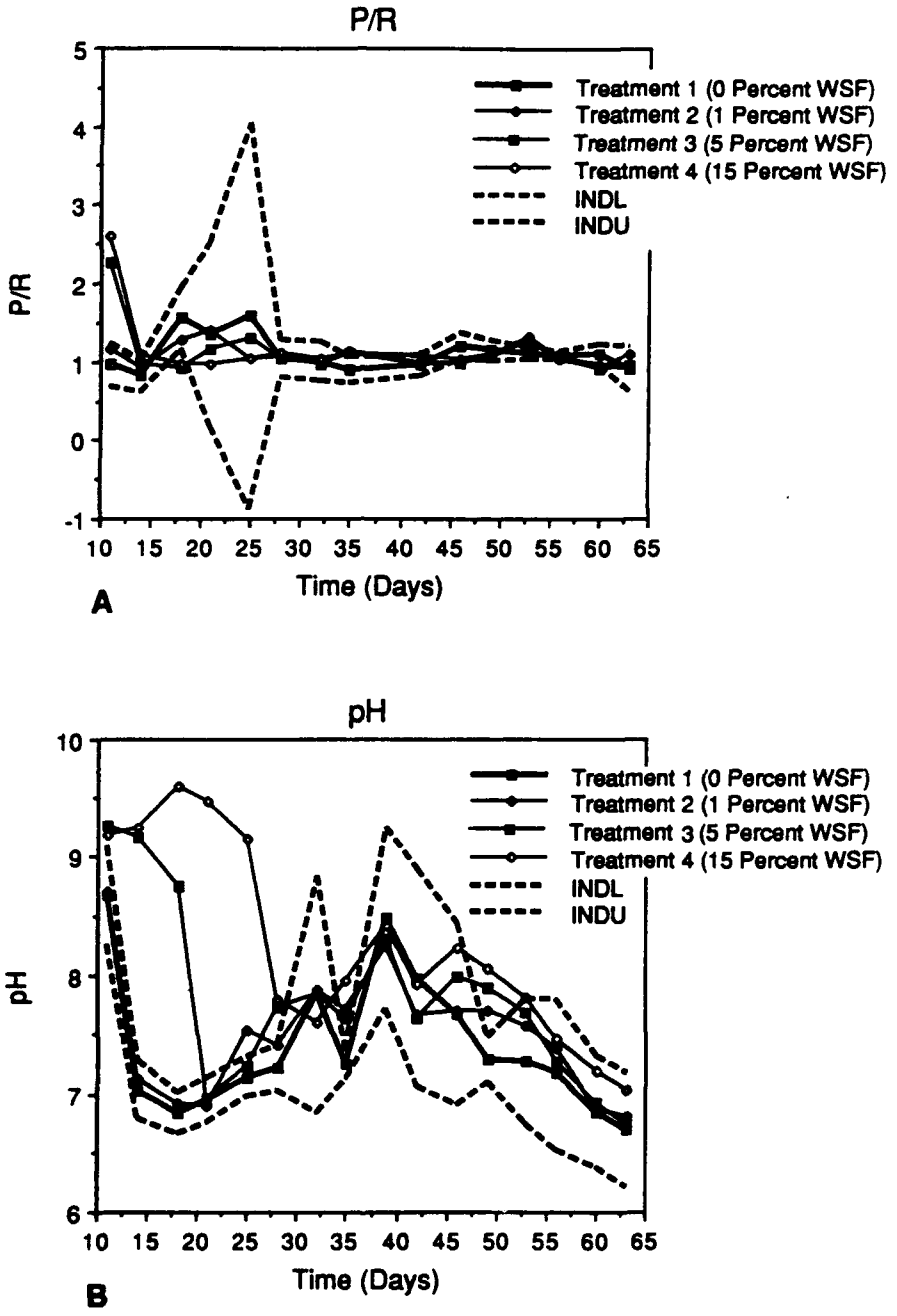


Fig. 7. Photosynthesis/respiration ratio and pH values from the SAM toxicity test of the WSF of Jet-A. Upper (INDU) and lower (INDL) limits of significance are shown as dashed lines.

3.4 *Multivariate results*

The significance levels for the three multivariate tests performed for each sampling day are graphed in Fig. 8. All three tests indicate that there were significant differences ($p \geq 0.95$) between treatment groups from day 11 through day 25, and again from day 46 through day 56. No consistent differences were observed from day 28 to day 39 and on days 60 and 63.

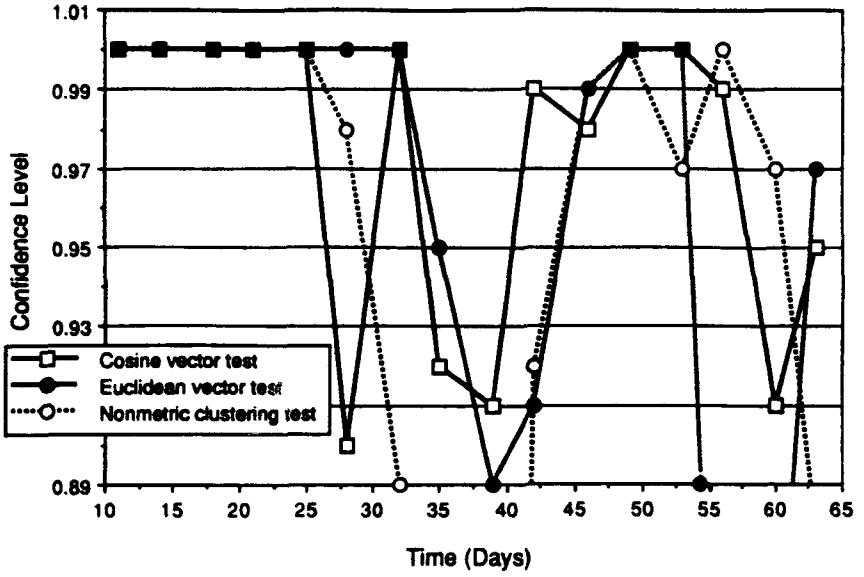


Fig. 8. Significance levels of three multivariate statistical tests (cosine vector, Euclidean vector, and nonmetric clustering) for the SAM toxicity test of the WSF of Jet-A.

In Fig. 9, the average cosine distances between the reference group and each of the three treatment groups are plotted on a log scale. The initial effect of the WSF dosing (day 11 to day 25) is apparent in the large distances between Treatment 1 and Treatment 4. Treatment 3 starts out distant from Treatment 1, but subsequently moves closer to the reference. The period of no significant differences (day 35 to day 46) is also obvious: none of the groups are especially far apart. During the second period of significant differences (day 46 to 56) a perfect dose/response relationship for all three treatments is seen, with higher doses becoming more distant from the control.

Using nonmetric clustering, we were able to list the variables that were the most important for separating the treatment group clusters for each day that measurements were collected (Table 3). This list shows that the specific variables that were most important for clustering changed over time. In addition, the number of variables used for clustering decreased from approximately 5–7 important variables on days 11–25 to ≤ 4 important variables from day 28 until the end of the experiment.

4. Discussion

Our examination of individual variables provided only a limited, and somewhat distorted view of the SAM response to Jet-A. The univariate data analysis did indeed show that there were some significant responses to the toxicant, especially during the first few weeks when the *Daphnia* populations declined and the algal

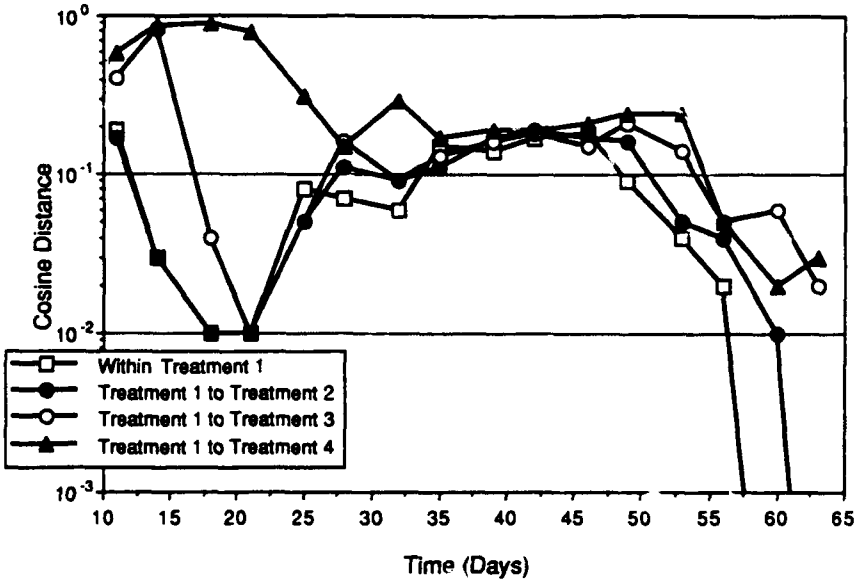


Fig. 9. Cosine distance from Treatment 1 to each of the remaining treatments for each sampling day. Smaller cosine distances indicate greater similarity between treatments.

populations peaked in the two highest treatment groups. However, the responses were scattered, and did not present a consistent pattern. Furthermore, the "significant" responses were actually gross aberrations of the microcosm, signifying wild

Table 3

Variables determined to be important in generating nonmetric clusters. Variables are listed in order of decreasing rank.

Day	*Important cluster variables (in rank order)*
11	M. *Daphnia, Chlorella, Chlamydomonas, Ulothrix*, S. *Daphnia, Selanastrum, Scenedesmus*
14	S. *Daphnia*, M. *Daphnia-Selenastrum*[1], *Chlamydomonas, Chlorella*, L. *Daphnia, Ankistrodesmus*
18	*Ankistrodesmus*, S. *Daphnia, Chlorella, Chlamydomonas, Selanastrum*, L. *Daphnia*
21	*Ankistrodesmus*, S. *Daphnia*, L. *Daphnia*-M. *Daphnia, Scenedesmus*
25	*Scenedesmus*, S. *Daphnia*, L. *Daphnia, Chlorella, Philodina*, M. *Daphnia*
28	*Ankistrodesmus*, L. *Daphnia, Scenedesmus*
32	S. *Daphnia*, M. *Daphnia, Ankistrodesmus, Chlorella*
35	*Ankistrodesmus*
39	M. *Daphnia-Selenastrum, Cyprinotus-Ankistrodesmus*
42	M. *Daphnia, Cyprinotus, Scenedesmus*
46	*Scenedesmus, Ankistrodesmus*, S. *Daphnia*, M. *Daphnia*
49	*Chlorella, Philodina, Ankistrodesmus, Lyngbya*
53	*Ankistrodesmus, Cyprinotus, Chlorella*
56	M. *Daphnia-Scenedesmus, Ankistrodesmus, Lyngbya*
60	*Lyngbya*, M. *Daphnia, Philodina, Chlorella*
63	*Chlorella, Ankistrodesmus, Philodina, Cyprinotus*

[1]Hyphen between variables denotes equal rank

swings in a taxon's population density. The confirmation of gross responses to a toxicant does not provide much more insight into the effects of the toxicant in an ecosystem than do short-term, single-species tests.

The multivariate statistics suggest a much more complex pattern of multiple divergences and convergences in the similarities between treatment groups. Much as an ecosystem could be expected to display the rise and fall of species assemblages, the SAMs appear to indicate that the first divergence was only the beginning of a series of responses.

The list of variables (Table 3) suggests that the first divergence, which occurred from about day 11 through day 32, resulted from predictable predator/prey interactions between *Daphnia* and algae. Theoretically, this divergence should be characterized by the following properties: (i) it should be fast, because the algae and *Daphnia* populations were introduced into the microcosm after being cultured in optimal laboratory conditions, in artificially high (and unstable) densities; (ii) it should be short-lived, because the populations are unstable in the nutrient-rich, early successional microcosm; (iii) there should be a tendency for the microcosms to drift away from their early treatment responses (especially because the WSF is essentially gone from the microcosms within a few days after its introduction) into more complex communities based on interactions between the remaining biotic constituents. This first divergence is the only type of response that is normally searched for in microcosm tests using conventional statistics, and is the response typically reported in SAM experiments.[9,10,32,33]

The second divergence occurred from about day 46 through day 60. During this time, other secondary consumers (e.g., *Cyprinotus* and *Philodina*) joined *Daphnia* and various algal taxa as being important in cluster development (see Table 3). The second divergence, therefore, may represent the long-term effects of the initial toxicant on a successionally more mature community. If so, the second divergence will be strongly influenced by detritus quality. Detritus is conditioned by bacteria and fungi, which are highly sensitive to toxins, but are not measured in the microcosm. Detritus that has passed through the gut of a consumer (e.g., *Daphnia*) is different from detritus that originates directly from unconsumed, dead algae. Therefore, the quality of the detritus may be highly affected by the treatment, but none of the factors influencing it are measured directly. Secondary consumers of detritus and bacteria (e.g., rotifers and ostracods) are no less affected by the quality of their food source than algal consumers, so the treatment-related alterations of the quality of detritus and bacteria will cause differences in the secondary consumer populations. Because this effect would occur late in the microcosm experiment and would be difficult to detect using univariate statistics, it would be easy to misinterpret as noise or as the effects of a degradation product.

Multiple divergences may also be explained without invoking direct impact of unseen biotic components of the system. The hypervolume defined by the multivariate data set for each treatment group may simply be moving in various directions and pass through the hypervolume of another treatment group at an in-

stant in time. When viewed during that time, the two groups would appear similar (or to have "recovered"). In reality, this similarity is only a momentary confluence.

Taken separately, none of the biotic variables measured in the SAM experiment could clearly identify the second divergence. Even pH, a variable with a low sampling error, did not consistently distinguish the second divergence. Without corroboration, the few pH values that fell outside the INDs late in the experiment would probably have been considered outliers. However, the three multivariate analyses demonstrated a clear, significant dose/response relationship for both the first and second divergences. Nonmetric clustering was also able to select the variables that were important in distinguishing the four treatment groups, although the variables contributing to the differentiation changed from sampling day to sampling day (Table 3). These data suggest that reliance upon any one variable (e.g., *Daphnia*, or an index of variables, probably would have missed the second divergence. The implications are important. Currently, only small sections of the ecosystems are monitored and a heavy reliance is placed upon so-called indicator species. Our data suggest that such a practice could produce misleading interpretations because the best indicator species will most likely change over the course of an experiment a season, or site, etc.

In summary, we found at least two divergences between the similarities of treatment groups for the WSF of Jet-A. Multivariate analyses were crucial in identifying these patterns; conventional univariate statistics provided only clues. Furthermore, the complexity of the multivariate responses showed that reliance upon any particular set of indicator species may be misleading in determining the effects of stressors upon biological communities.

Acknowledgements

This research is supported by USAFOSR Grant No. AFOSR-91-2091 DEF. Robert V. Sahakian and Lester W. Keel participated in the preparation and sampling of the microcosms. Mark V. Haley, U.S. Army Chemical Research Division and Engineering Center kindly supplied the original cultures used in these experiments.

References

1 ASTM E1366-91. (1991): Standard practice for standardized aquatic microcosm: fresh water, volume 11.04. American Society for Testing and Materials, Philadelphia, PA.

2 Taub, F. B. (1969): Genotobiotic models of freshwater communities. *Verh. Internat. Verein. Limnol.* **17**: 485–496.

3 Taub, F. B. (1976): Demonstration of pollution effects in aquatic microcosms. *Intern. J. Environ. Stud.* **10**: 23–33.

4 Taub, F. B. and Crow, M. E. (1978): Loss of a critical species in a model (laboratory) ecosystem. *Verh. Internat. Verein. Limnol.* **20**: 1270–1276.

5 Crow, M. E. and Taub, F. B. (1979): Designing a microcosm bioassay to detect ecosystem level effects. *Intern. J. Environ. Stud.* **13**: 141–147.

6 Taub, F. B., Crow, M. E. and Hartmann, H. J. (1980): Responses of aquatic microcosms to acute mortality. In J. P. Giesy, Jr., Ed.: *Microcosms in Ecological Research*, Technical Information Center, U. S. Department of Energy, Washington, D.C., pp. 513–535.

7 Kindig, A. C., Loveday, L. C. and Taub., F. B. (1983): Differential sensitivity of new versus mature synthetic microcosms to streptomycin sulfate treatment. In W. E. Bishop *et al.*, Eds.: *Aquatic Toxicology and Hazard Assessment*: Sixth Symposium, ASTM 802, American Society for Testing and Materials, Philadelphia, PA, pp. 192–203.

8 Taub, F. B. and Read, P. L. (1983): Standardized Aquatic Microcosm Protocol: Final Report Contract No. 223-80-2352. Vol II. Food and Drug Administration, Washington, D.C.

9 Taub, F. B., Kinding, A. C. and Conquest, L. L. (1987): Interlaboratory testing of a standardized aquatic microcosm. In: W. J.. Adams *et al.*, Eds.: *Aquatic Toxicology and Hazard Assessment*: 10th Volume, ASTM STP 971, American Society for Testing and Materials, Philadelphia, PA, pp. 385–405.

10 Taub, F. B., Kindig, A. C., Conquest, L. L. and Meador, J. P. (1988): Results of the interlaboratory testing of the standardized aquatic microcosm protocol. In G. Suter and M. Lewis, Eds.: *Aquatic Toxicology and Hazard Assessment*: Eleventh Symposium ASTM. American Society for Testing and Materials, Philadelphia, PA,

11 Taub, F. B. (1988): Standardized aquatic microcosm-development and testing. In A. Boudou and F. Ribeyre, Eds.: *Aquatic Ecotoxicology*, Vol. II, CRC Press, Boca Raton, FL, pp. 47–92.

12 Taub, F. B. (1989): Standardized aquatic microcosms. *Environm. Sci. Technol.* **23(9)**: 1064–1066.

13 Conquest, L. L. and Taub, F. B. (1989): Repeatability and reproducibility of the standard aquatic microcosm: Statistical properties. In U. M. Cowgill and L. R. Williams, Eds.: *Aquatic Toxicology and Hazard Assessment*: 12th Volume, ASTM STP 1027, American Society for Testing and Materials, Philadelphia, PA, pp. 159–177.

14 Kersting, K. (1984): Development and use of an aquatic micro-ecosystem as a test system for toxic substances. Properties of an aquatic micro-ecosystem IV. *Int. Revue ges. Hydrobiol.* **69(4)**: 567–607.

15 Kersting, K. (1985): Properties of an aquatic micro-ecosystem. V. Ten years of observations of the prototype. *Verh. Internat. Verein. Limnol.* **22**: 3040–3045.

16 Kersting, K. (1988): Normalized ecosystem strain in micro-ecosystems using different sets of state variables. *Verh. Internat. Verin. Limnol.* **23**: 1641–1646.

17 Johnson, A. R. (1988): Diagnostic variables as predictors of ecological risk. *Environ. Manag.* **12(4)**: 515–523.

18 Johnson, A. R. (1988): Evaluating ecosystem response to toxicant stress: a state space approach. In W. J. Adams *et al.* Eds.: *Aquatic Toxicology and Hazard Assessment*: 10th Volume, ASTM STP 971, American Society for Testing and Materials, Philadelphia, PA, pp. 275–285.

19 Matthews, G. B. and Matthews, R. A. (1991): A model for describing community change. In: Pesticides in Natural Systems: How Can Their Effects Be Monitored? *Proceedings of the Conference, Environmental Research Laboratory*/ORD, Corvalis, OR, EPA 9109/9-91011.

20 Matthews, G. B., Matthews, R. A. and Hachmöller, B. (1991): Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Can. J. Fish: Aquat. Sci.* **48(11)**: 2184–2190.

21 Matthews, R. A., Matthews, G. B. and Ehinger, W. J. (1991): Classification and ordination of limnological data: a comparison of analytical tools. *Ecol. Model.* **53**: 167–187.

22 Matthews, G. B. and Hearne, J. W. (1991): Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13(2)**: 175–184.

23 ASTM D3710. (1986): Boiling range distribution of gasoline and gasoline fractions by gas chromatography. *Annual Book of ASTM Standards* 5.03: 99–113.

24 ASTM D2887. (1986): Boiling range distribution of petroleum fractions by gas chromatography. *Annual Book of ASTM Standards*, 5.02: 658–667.

25 ASTM D-2 Proposal (P167). (1989): Manual on hydrocarbon analysis 4th Ed. Series MNL3. American Society for Testing and Materials, Philadelphia, PA.

26 ASTM E729-88a. (1988): Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphibians, Volume 11.04 American Society for Testing and Materials, Philadelphia, PA.
27 Good, I. J. (1982): An index of separateness of clusters and a permutation test for its significance. *J. Statist. Comp. Simul.* **15**: 81–84.
28 Smith, E. P., Pontasch, K. W. and Cairns, J., Jr. (1990): Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.* **24(4)**: 507–514.
29 Noreen, E. W. (1989): Computer Intensive Methods for Testing Hypotheses. Wiley-Interscience, New York, NY.
30 Fienberg, S. E. (1985): The Analysis of Cross-Classified Categorical Data. MIT Press, Cambridge, MA.
31 Press, W. H., Flannery, B. P., Teukolsky, A. A. and Vetterline, W. T. (1990): *Numerical Recipes in C, the Art of Scientific Computing*. Cambridge University Press, New York, NY.
32 Landis, W. G., Chester, N. A., Haley, M. V., Johnson, D. W., Muse, W. T., Jr. and Tauber, R. M. (1989): The utility of the standard aquatic microcosm as a standard method for ecotoxicological evaluation. In G. Suter and M. Adams, Eds.: *Aquatic Toxicology and Environmental Fate*: Eleventh Volume ASTM STP-1007, American Society for Testing and Materials, Philadelphia, PA, pp. 353–367.
33 Landis, W. R., Haley, M. V. and Chester, N. A. (In press): The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In W.G. Landis, J. Hughes and M. Lewis, Eds.: Environmental Toxicology and Risk Assessment: First Volume ASTM STP-1179. American Society for Testing and Materials, Philadelphia, PA.

Ecotoxicology **2**. 271–300 (1993)

Multivariate analysis of the impacts of the turbine fuel JP-4 in a microcosm toxicity test with implications for the evaluation of ecosystem dynamics and risk assessment

WAYNE G. LANDIS[1]*, ROBIN A. MATTHEWS[1], APRIL J. MARKIEWICZ[1] and GEOFFREY B. MATTHEWS[2]

[1]*Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA 98225 USA*
[2]*Computer Science Department, Western Washington University, Bellingham, WA 98225 USA*

Received 7 April 1993; revised and accepted 22 July 1993

Turbine fuels are often the only aviation fuel available in most of the world. Turbine fuels consist of numerous constituents with varying water solubilities, volatilities and toxicities. This study investigates the toxicity of the water soluble fraction (WSF) of JP-4 using the Standard Aquatic Microcosm (SAM). Multivariate analysis of the complex data, including the relatively new method of nonmetric clustering, was used and compared to more traditional analyses. Particular emphasis is placed on ecosystem dynamics in multivariate space.

The WSF is prepared by vigorously mixing the fuel and the SAM microcosm media in a separatory funnel. The water phase, which contains the water-soluble fraction of JP-4 is then collected. The SAM experiment was conducted using concentrations of 0.0, 1.5 and 15% WSF. The WSF is added on day 7 of the experiments by removing 450 ml from each microcosm including the controls, then adding the appropriate amount of toxicant solution and finally bringing the final volume to 3 L with microcosm media. Analysis of the WSF was performed by purge and trap gas chromatography. The organic constituents of the WSF were not recoverable from the water column within several days of the addition of the toxicant. However, the impact of the WSF on the microcosm was apparent. In the highest initial concentration treatment group an algal bloom ensued, generated by the apparent toxicity of the WSF of JP-4 to the daphnids. As the daphnid populations recovered the algal populations decreased to control values. Multivariate methods clearly demonstrated this initial impact along with an additional oscillation seperating the four treatment groups in the latter segment of the experiment. Apparent recovery may be an artifact of the projections used to describe the multivariate data. The variables that were most important in distinguishing the four groups shifted during the course of the 63 day experiment. Even this simple microcosm exhibited a variety of dynamics, with implications for biomonitoring schemes and ecological risk assessments.

Keywords: jet fuel; microcosm; multivariate statistics; nonmetric clustering; risk assessment.

Introduction

As this is written, the United States Environmental Protection Agency has suspended the requirement for conducting ecosystem level studies for pesticide registration (Fisher

*To whom correspondence should be addressed.

1992). Although many factors contributed to the action, apparently the field and pond mesocosm tests that were conducted did not contribute to the evaluation of risk of pesticides in a timely and cost effective manner.

Over the last 15 years a variety of multispecies toxicity tests have been developed with the hope that in doing so, the increased complexity of the test would result in more realistic, community-level responses to the toxicant. However, the addition of more than one species, and the generally longer time periods associated with these multispecies tests, also result in much more complex data sets. Distinguishing toxicant effects from other community-level changes has become one of the most critical obstacles to the interpretation of multispecies data sets.

Multispecies toxicity tests are usually referred to as microcosms or mesocosms, although a clear definition of the size or complexity to distinguish these terms has not been put forth. Multispecies toxicity tests range from approximately 1 L (e.g., mixed flask cultures) to thousands of litres, as in the case of the pond mesocosms used in pesticide registration testing. The number of species and origin of those taxa can vary widely. In the Standardized Aquatic Microcosm (SAM) developed by Taub and colleagues (Taub 1969, 1976, Taub and Crow 1978, Crow and Taub 1979, Taub *et al.* 1980, Kindig *et al.* 1983, Taub *et al.* 1987, Taub *et al.* 1988, Taub 1988, 1989, Conquest and Taub 1989) the physical, chemical, and biological components are defined as to species, media and substrate (see Table 1 and Fig. 1). In other systems colonization by the importation of sediment or by repeated inoculation from a natural source is used to establish the model system. Larger systems often use a combination of means to start and maintain a multispecies, interactive community.

One of the major difficulties in the evaluation of multispecies toxicity tests has been the difficulty in the analysis of the large data set on a level consistent with the goals of the toxicity test. Typically, the goals of the toxicity test are:

(1) to detect changes in the population dynamics of the individual taxa that would not be apparent in single species tests; and,

(2) to detect community-level differences that are correlated with treatment groups thereby representing a deviation from the control group.

A number of methods have been developed to attempt to satisfy the goals of multispecies toxicity testing. Analysis of variance (ANOVA) is the classical method to examine single variable differences from the control group. However, because multispecies toxicity tests generally run for weeks or even months, there are problems with using conventional ANOVA. These include the increasing likelihood of introducing a Type II error (accepting a false null-hypothesis), temporal dependence of the variables, and the difficulty of graphically representing the data set. Conquest and Taub (1989) developed a method to overcome some of the problems by using intervals of non-significant difference (IND). This method corrects for the likelihood of Type II errors and produces intervals that are easily graphed to ease examination. The method is routinely used to examine data from SAM toxicity tests, and it is applicable to other multivariate toxicity tests. The major drawback is the examination of a single variable at a time over the course of the experiment. While this addresses the first goal in multispecies toxicity testing, listed above, it ignores the second. In many instances, community-level responses are not as straightforward as the classical predator/prey or nutrient limitation dynamics usually picked as examples of single-species responses that represent complex interactions.

Table 1. Summary of test conditions for conducting SAM JP-4

Organisms	
Organisms per chamber:	Algae (added on Day O at initial concentration of 10^3 cells for each algae species: *Anabaena cylindrica*, *Ankistrodesmus* sp., *Chlamydomonas reinhardi* 90, *Chlorella vulgaris*, *Lyngbya* sp., *Scenedesmus obliquus*, *Selenastrum capricornutum*, *Stigeoclonium* sp., and *Ulothrix* sp. Animals (added on Day 4 at the initial numbers indicated in parentheses): *Daphnia magna* (16 per microcosm), *Cypridopsis* sp. (ostracod) (6 per microcosm), *Tetrahymena thermophila* [protozoa] (0.1 per mL), and *Philodina* sp. (rotifer) (0.03 per mL)
Experimental design	
Test vessel type and size:	One-gallon (3.8 L) glass jars 16.0 cm wide at the shoulder, 25 cm tall with 10.6 cm openings
Medium volume:	3 L added to each container
Number of replicates × concentrations:	6 × 4
Reinoculation:	Once per week add one drop (ca 0.05 mL) to each microcosm from a mix of the ten species = 5×10^2 cells of each alga added per microcosm
Addition of test materials:	Test material added day 7 by removing 450 mL from each container and then adding appropriate amounts of the WSF to produce concentrations of 0, 1, 5 and 15% WSF. After toxicant addition the final volume was adjusted to 3 L.
Sampling frequency:	2 times each week
Test duration:	63 days
Physical and chemical parameters	
Temperature:	20 to 25 °C
Light intensity:	80 μE m^{-2} photosynthetically active radiation s^{-1} (850 to 1000 fc)
Photoperiod:	12 h light/12 h dark
Medium:	Medium T82MV
Sediment:	Composed of silica sand (200 g), ground, crude chitin (0.5 g), and cellulose powder (0.5 g) added to each container
Measurements:	Algal, invertebrate and protozoa counts, pH, dissolved oxygen, optical density. Parameters calculated included the concentrations of each of the species, DO, DO gain and loss, net photosynthesis/respiration ratio (P/R), pH, algal species diversity, daphnid fecundity, algal biovolume, and biovolume of available algae.

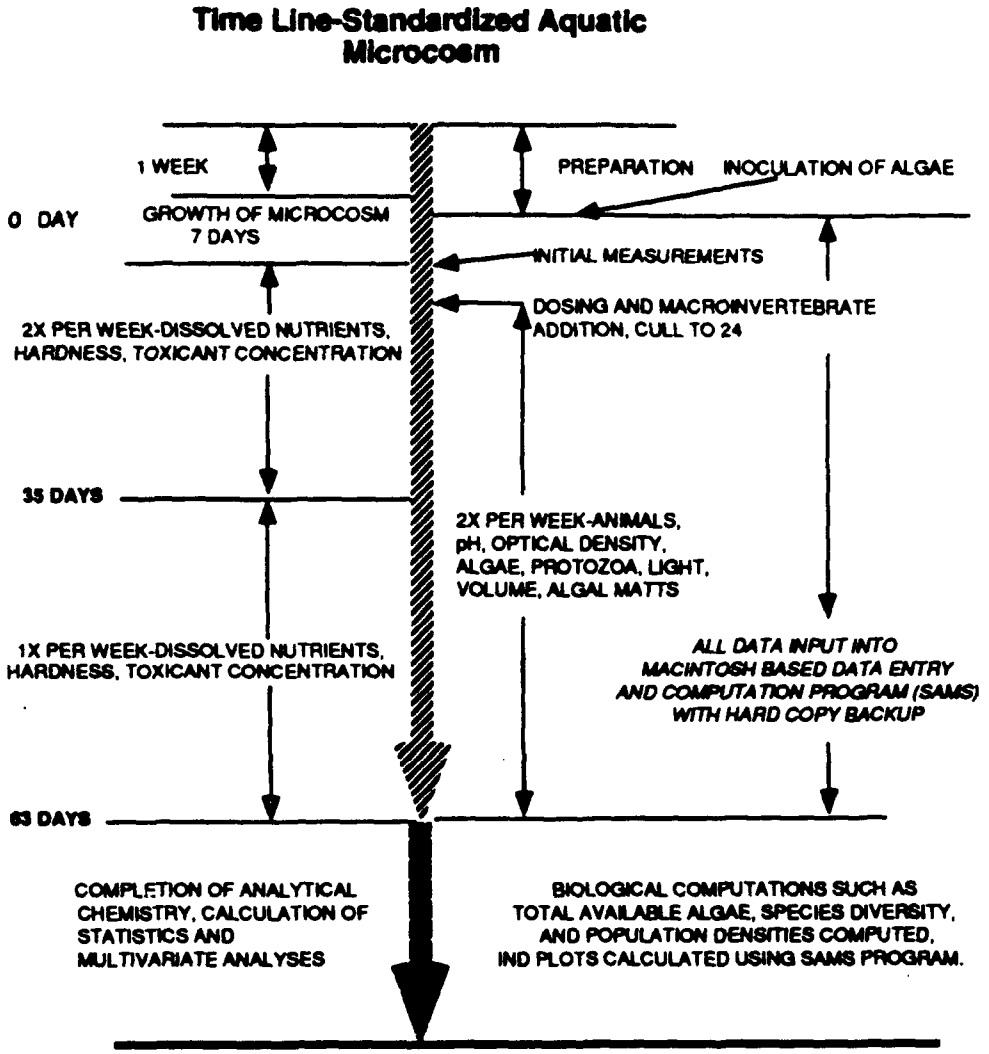


Fig. 1. Timeline for the standardized aquatic microcosm JP-4 experiment. Each step of this 63 day protocol is choreographed according to ASTM E 1366–91. The modifications to the protocol are the elimination of *Nitchia*, *Hyalella azteca*, modification of the method for toxicant delivery and the substitution of *T. thermophila* BIV for the hypotrichous ciliate.

Multivariate methods have proved promising as a method of incorporating all of the dimensions of an ecosystem. One of the first methods used in toxicity testing was the calculation of ecosystem strain developed by Kersting (1984, 1985, 1988) for a relatively simple (three species) microcosm. This method has the advantage of using all of the measured parameters of an ecosystem to look for treatment-related differences. At about the same time, Johnson (1988a, b) developed a multivariate algorithm using the n-dimensional coordinates of a multivariate data set and the distances between these coordinates as a measure of divergence between treatment groups. Both of these

methods have the advantage of examining the ecosystem as a whole rather than by single variables, and can track such processes as succession, recovery and the deviation of a system due to an anthropogenic input.

However, a major disadvantage of both these methods, and of many conventional multivariate methods, is that all of the data are often incorporated without regard to the units of measurement or the appropriateness of including all variables in the analysis. It can be difficult to combine variables such as pH, with units ranging from 0–14, with the numbers of bacterial cells per ml, where low numbers are in the 10^6 range, to say nothing of the conceptual difficulties of adding pH units to counts. Similarly, random variables (i.e., variables with no treatment-related response) indiscriminately incorporated into the analysis may contribute so much noise that they overshadow variables that do show treatment-related effects.

Ideally, a multivariate statistical test used for evaluating complex data sets will have the following characteristics:

(1) It will not combine counts from dissimilar taxa by means of sums of squares, or other *ad hoc* mathematical techniques, as in the Euclidean and cosine distance measures.

(2) It will not require transformations of the data, such as normalizing the variance.

(3) It will work without modification on incomplete data sets.

(4) It will work without further assumptions on different data types (e.g., species counts or presence/absence data).

(5) Significance of a taxon to the analysis will not be dependent on the absolute size of its count, so that taxa having a small total variance, such as rare taxa, can compete in importance with common taxa, and taxa with a large, random variance will not automatically be selected, to the exclusion of others.

(6) It will provide an integral measure of 'how good' the analysis is, i.e. whether the data set differs from a random collection of points.

(7) It will, in some cases, identify a subset of the taxa that serve as reliable indicators of the physical environment.

Recently developed for the analysis of ecological data, nonmetric clustering is a multivariate derivative of artificial intelligence research that satisfies all these criteria, and has the potential of circumventing many of the problems of conventional multivariate analysis.

In this paper, we use ANOVA and intervals of non-significant difference, and three multivariate techniques to search for meaningful patterns in the data set from a SAM toxicity test using Jet-A turbine fuel. The multivariate techniques include two conventional tests based on the ratio of multivariate metric distances (Euclidean distance and cosine of the vector distance), and one relatively new program, RIFFLE, which employs nonmetric clustering and association analysis (Matthews and Hearne 1991). All three of the multivariate techniques have proven useful in analysing complex ecological data sets (Matthews *et al.* 1991a, b). Of the three, only nonmetric clustering meets all of the criteria listed above (Matthews and Matthews 1991). The major disadvantage of the RIFFLE program is that, in order to find a clustering of the data points with the desirable qualities listed above, a massive search through thousands of potential clustering candidates is made before settling on the 'right' one. Even after this search, there is no guarantee that RIFFLE finds an optimal clustering. However, in our experience, RIFFLE does find an excellent clustering in reasonable time.

Jet fuels or perhaps more accurately, turbine fuels, are one of the primary fuels for internal combustion engines worldwide and certainly are the most widely available aviation fuel. Over the last 15 years virtually all of the commercial airline operations and charter operations have converted to a turbine engine because of the inherent low operating cost of the power plant, its reliability, and in part to the availability of fuel even in underdeveloped areas. In the US military there has been a progressive replacement of conventional piston engine vehicles with turbine equivalents. Standardization on a single type of turbine fuel to relieve logistical demands is also underway. Given the overwhelming predominance of turbine fuel, a fuel spill or accidental release of aviation fuel will probably be one of the prevalent turbine fuels: Jet-A, used for commercial and general aviation; JP-4, the standard fuel of the US Air Force and Army Aviation; and JP-5, the naval equivalent of JP-4. JP-8 is a new fuel proposed as the standard for all military vehicles using turbine engines.

Along with the environmental considerations, turbine fuels also offer advantages as model complex toxicants for toxicological research. Because of their use as aviation fuel, turbine fuels are produced to stringent specifications designed to ensure the safety of flight. Therefore, the overall general properties of these materials are tightly controlled. In addition, standard archived samples of the military fuels are maintained for toxicological studies at Wright Patterson, AFB. Jet fuels also tend to be less explosive and less volatile than gasoline, making the materials easier and safer to use. These properties make jet fuels an ideal material for the investigation of the effects of complex mixtures upon community dynamics. Like all petroleum products, however, the exact identity of the constituents varies according to the original crude and the refining process.

This paper reports the effects of low concentration of the water soluble fraction of JP-4 on the community incorporated in the SAM. The effects of the WSF on the microcosm communities were subtle. An early increase in algal density was apparent in the treatment groups containing the highest concentrations of the WSF and was matched by a decrease in daphnid populations. Multivariate analysis proved to be more powerful and efficient in highlighting important variables and processes than ANOVA. The variables that were most important were those distinguishing where treatment-related effects shifted during the course of the experiment. The multivariate analysis also detected oscillations in the similarity of the control and dosed groups that were not apparent using conventional univariate tests. The oscillations may be due to the inherent perturbations in community dynamics and interactions, or the effects upon the segments of the community not directly measured, the bacterial detritivores. We also discuss the implications of this research with regards to the use of indices and the conduct of environmental risk assessments.

Materials and methods

Reagents

All chemicals used in the culture of the organisms and in the formulation of the microcosm media were reagent grade or as specified by the ASTM method.

JP-4 was supplied by the US Air Force Toxicology Laboratory at Wright Patterson, AFB, Ohio.

Water soluble fraction (WSF)

The WSF of JP-4 was prepared in glassware washed in nonphosphate soap, rinsed, then soaked in 2N HCl for at least 1h, rinsed ten times with distilled water, dried and finally autoclaved for 30 min. Microcosm medium, T82MV, acted as the diluent for the water fraction of the WSF.

Twenty five ml of JP-4 is added to the 1 l separatory funnel containing 1 l of T82MV, and is agitated as follows:
(1) Shake separatory funnel for 5 min, releasing built up pressure as necessary, (2) allow funnel contents to remain undisturbed for 15 min, (3) shake contents for 5 min, allow to stand 15 min, (4) continue same pattern for a total time of 1 h, and finally (5) allow separatory funnel contents to remain undisturbed for 8 h. At the end of this procedure the mixture was allowed to stand overnight. The next day all but 100 ml of T82MV/WSF of jet fuel mixture from the separatory funnel (leaving the lighter, insoluble fuel mixture in the flask) was drained into a cleaned, sterile 1 l amber glass bottle and capped with a Teflon-lined screw cap. The WSF was used within 24 h or stored at 4 °C for no longer than 48 h before use as toxicant mixture.

Gas chromatography of WSF

This protocol utilizes a Tekmar LSC 2000 Purge and Trap (P&T) concentrator system in tandem with a Hewlett Packard 5890A Gas Chromatograph with a Flame Ionization Detector (FID) (ASTM D3710 1988: ASTM D2887 1988: Westendorf 1986). Instrument blanks and deionized distilled water blanks are used to verify the cleanliness of the P&T and GC columns prior to analysis of samples. A 5 ml sample is injected into a 5 ml sparger, purged with pre-purified nitrogen gas for 11 min and dry purged for 4 min. Volatile hydrocarbons, purged from the sample and collected on the Tenax/Silica Gel column, are desorbed at 180 °C directly onto the gas chromatograph SPB-5, 30 m × 0.53 mm ID 1.5 μm film, fused silica capillary column. The column, at 35 °C, is held at that temperature for 2 min, increased to 225 °C at 12 °C min^{-1} and held at that temperature for 5 min. A Spectra-Physics 4290 Integrator records the FID signal output of the volatile hydrocarbons that have been separated and eluted from the column by molecular weight.

Identification and quantification of GC fractions

Qualitative identification of some components in the WSF of the JP-4 fuel used as the toxicant in the microcosm test, were determined using a Simulated Distillation (SIMDIS) Calibration Mixture. The ASTM Method D3710 Qualitative Calibration Mixture is the standard test method for determining the Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography. This mixture was used as a calibration standard to determine the retention times for each known component in the mixture against which unknown components, in the WSF of the Jet fuel mixture, were compared and identified.

Quantitative estimates of some components of the WSF were made by comparing sample chromatographs to certified n-paraffin and n-naphtha chromatograph standards, prepared and analysed under the same P&T/GC conditions.

Algal toxicity tests

In order to estimate the relative toxicities of the JP-4 mixture and to set the concentrations for the microcosm, a series of short-term toxicity tests were performed (ASTM E 1218 1991). Algal growth inhibition tests were performed using *Ankistrodesmus falcatus* and *Selenastrum capricornutum* strains identical to those used in the SAM toxicity tests.

Test algae were grown in a semi-flow through culture apparatus on the microcosm media T82MV and taken during log phase growth for inoculation into the test flasks. Erlenmeyer flasks (250 ml) were used as test chambers, with serial dilutions of the WSF at concentrations of 0.0, 6.25, 12.5, 25, 50 and 100% placed in the flasks. The test organisms were added at a concentration of approximately 3.0×10^4 cells ml^{-1}. Total volume was 100 ml with two replicates of controls and the test concentrations used. Test mixtures were incubated at 20.0 °C ± 1.0 °C with a 12:12 h light/dark cycle. Using a Newbauer Counting Chamber, cell densities were determined every 24 h for the 96 h duration of the test.

The cell numbers were then plotted against the concentrations. If possible, a least square regression line was drawn and the LC_{50} (the concentration at which algal growth is inhibited to 50% of the control) determined. ANOVA was then run on the replicates to determine if any of the groups are significantly different.

SAM protocol

The 64-day SAM protocol has been described previously (ASTM E 1366–91 1991). Table 1 describes the organisms, conditions and modifications of ASTM E1366–91 for this particular experiment. Briefly, the microcosms were prepared by the introduction of ten algal, four invertebrate, and one bacterial species into 3 l of sterile defined medium. Test containers were 4 l glass jars. An autoclaved sediment consisting of 200 g silica sand and 0.5 g of ground chitin is autoclaved in the experimental jar immersed in a water bath to a point above the sand and chitin level during sterilization. This procedure helps prevent breakage of the jars and subsequent loss of replicates.

Numbers of organisms, dissolved oxygen (DO) and pH were determined twice weekly. Room temperature was 20 °C ± 2°. Illumination was 79.2 μEm^{-2} photosynthetically active radiation s^{-1} with a range of 78.6–80.4 and a 16.8 day/night cycle.

Two major modifications were made to the SAM protocol. The first was the means of toxicant delivery. Test material was added on day 7 by stirring each microcosm, removing 450 ml from each container and then adding appropriate amounts of the WSF to produce concentrations of 0, 1, 5 and 15 % WSF. After toxicant addition the final volume was adjusted to 3l. No attempt was made to filter and retain the organisms withdrawn during the removal of the 450 ml prior to toxicant addition. All graphs and statistical analysis start with the first sampling day, day 11.

The second modification was the substitution of *Tetrahymena thermophila* BIV for the hypotrichous ciliate used in past experiments. The hypotrichous ciliate was becoming increasingly difficult to culture, probably due to the age of the clone. *T. thermophila* has routinely been used in biochemical research and in detoxification studies of organophosphates (Landis *et al.* 1985, 1987, 1991). Using SAM controls, constructed prior to this experiment, it was demonstrated that the *T. thermophila* populations were able to exist within the system. *T. thermophila* are maintained steriley in a 3% proteous peptone distilled water media at 20 °C with routine biweekly transfers to perpetuate the stocks.

The results presented below demonstrate the suitability of the *Tetrahymena* for inclusion in the protocol.

Data analysis

All data were recorded onto standard computer entry forms and checked for accuracy. The data was then keyed into the SAMS data analysis program and checked for accuracy. Parameters calculated included the concentrations of each of the species, DO, DO gain and loss, net photosynthesis/respiration ratio (P/R), pH, algal species diversity, algal biovolume, and biovolume of available algae. The statistical significance of these parameters compared to the controls was also computed for each sampling day using the Interval of Non-significant Difference (IND) plots developed by Conquest. Note that algal biovolume, algal species diversity and available algae are all derived variables based on the algal counts. The net photosynthesis/respiration ratio is not derived using ^{14}C methods but by comparing oxygen concentrations before lights on, at the end of the photosynthetic period, and then at the next morning, as specified in the standard protocol. Photosynthesis/respiration ratio was the variable used during the analysis to incorporate these measurements.

The multivariate methods used in the analysis include cosine and vector distances and nonmetric clustering. All of these methods have been previously described (Matthews *et al.* 1991, Landis *et al.* 1993a, b) and are reviewed in Appendix A Table 2 lists the variables used in the clustering process.

Results

Algal toxicity results

The WSF of JP-4 was not very toxic on a percentage (v/v) of the total culture media. Effects were compared by computing the area underneath the growth curve for both the 96 h experiments. As determined by graphical analysis, since 100% inhibition compared to controls was not achieved, the IC_{50} for *Ankistrodesmus* was 57% WSF and for *Selenastrum* 95% WSF.

Persistence of the JP-4 WSF

Seven compounds, benzene, 2,4 dimthylpentane, ethylbenzene, 2-methylpentane, 2-methylpropane, o-xylene and toluene, were tracked using GC analysis during the course of the SAM experiment. Figure 2 is an area graph that presents both the concentrations of the individual components along with the totals of these seven materials in microcosms of Treatment 4. As can be readily seen, 504 h after dosing, the relative concentrations of these materials have rapidly disappeared. After week three, only 2-methylpentane and 2-methylpropane are detectable. Since only the 2-methylpropane is present 672 h after dosing, this material may be the final biodegradative product of the absorbed fraction of the WSF, and is being investigated in more detail

Patterns in algal communities

The largest increase in algal population density occurred in treatment 4 (Fig. 3). The peak density is approximately twice that of the control replicates at day 21. After the initial bloom in treatment 4, no particular dose-related pattern is discernible. *Lyngbya* makes up a substantial portion of the algal community in each treatment group, which

Table 2. Biotic parameters used in the multivariate statistical tests. Biotic variables such as diversity, available biovolume, and total algal biovolume are not used since they are derived from the variables listed below. Including derived variables weights some parameters more than others since some like *Anabaena* can be used alone and again in the calculation of total algal biovolume

Biotic parameter
Anabaena
Ankistrodesmus
Chlamydomonas
Chlorella
Daphnia
Ephipia
Small Daphnia
Medium Daphnia
Large Daphnia
Tetrahymena
Lyngbya
Miscellaneous sp.
Ostracod (*Cyprinotus*)
Philodina (Rotifer)
Scenedesmus
Selanastrum
Stigeoclonium
Ulothrix

is historically unusual. The number of algal species, as enumerated by the counting technique, also generally declines in each of the treatment groups, but in a general sense not related to dose.

Daphnid populations.

Each of the treatment groups exhibited similar dynamics (Fig. 4). None of the groups were statistically different from the control groups using conventional analysis of variance approaches. Minor perturbations in the timing of the peaks may have occurred, but by day 50 the means of each group were very similar.

Ostracod populations.

At the end of the experiment, the average population density in the control treatments is approximately twice that of treatment 4, the highest toxicant concentration (Fig. 5). Population density in the two treatment groups with the highest toxicant concentrations, decline below the no dose treatment and the lowest treatment densities. This pattern is apparent graphically from day 53 onward. Conventional analysis such as the IND plot

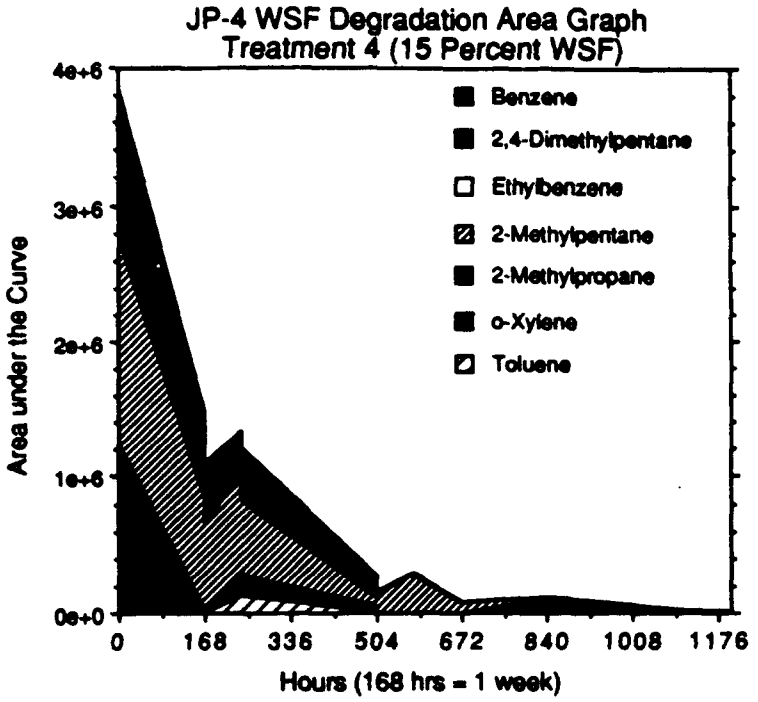


Fig. 2. Purge and trap Gas Chromatography results for the WSF of JP-4. A substantial reduction in the number and concentration of the WSF constituents is apparent two weeks after dosing in Treatment 4. At the end of the SAM experiment the fractions are at relatively low concentrations.

does not pick any date as significantly different from the control. The probability of the order remaining consistent on five consecutive dates when derived from a common population by chance alone and assuming independence of each group is small ($12^* (1/4!)^5 = 0.0000015$).

Philodina *and* Tetrahymena *populations.*

Tetrahymena survived in each of the treatment groups until near the end of the experiment (Fig. 6a). No specific dose related pattern was apparent although a two sampling period bloom (days 25 and 27) was apparent for Treatment 2. Unfortunately the error in sampling and the inherent asynchrony in Protistan reproduction prevented the result from being detectable using conventional methods. *Philodina* did not appear in appreciable numbers until after day 25 in any of the treatments. Day 53 showed a dramatic increase in treatments 3 and 4 followed by a decline, so that by day 60 all treatments were similar. Although suggestive, the results are not significant; the large overlap of the standard deviation apparent (Fig. 6b). The difficulty in sampling rapidly growing and declining populations in asynchronous growth is apparent. Although trends may be suggested, conventional analysis does not detect a significant effect.

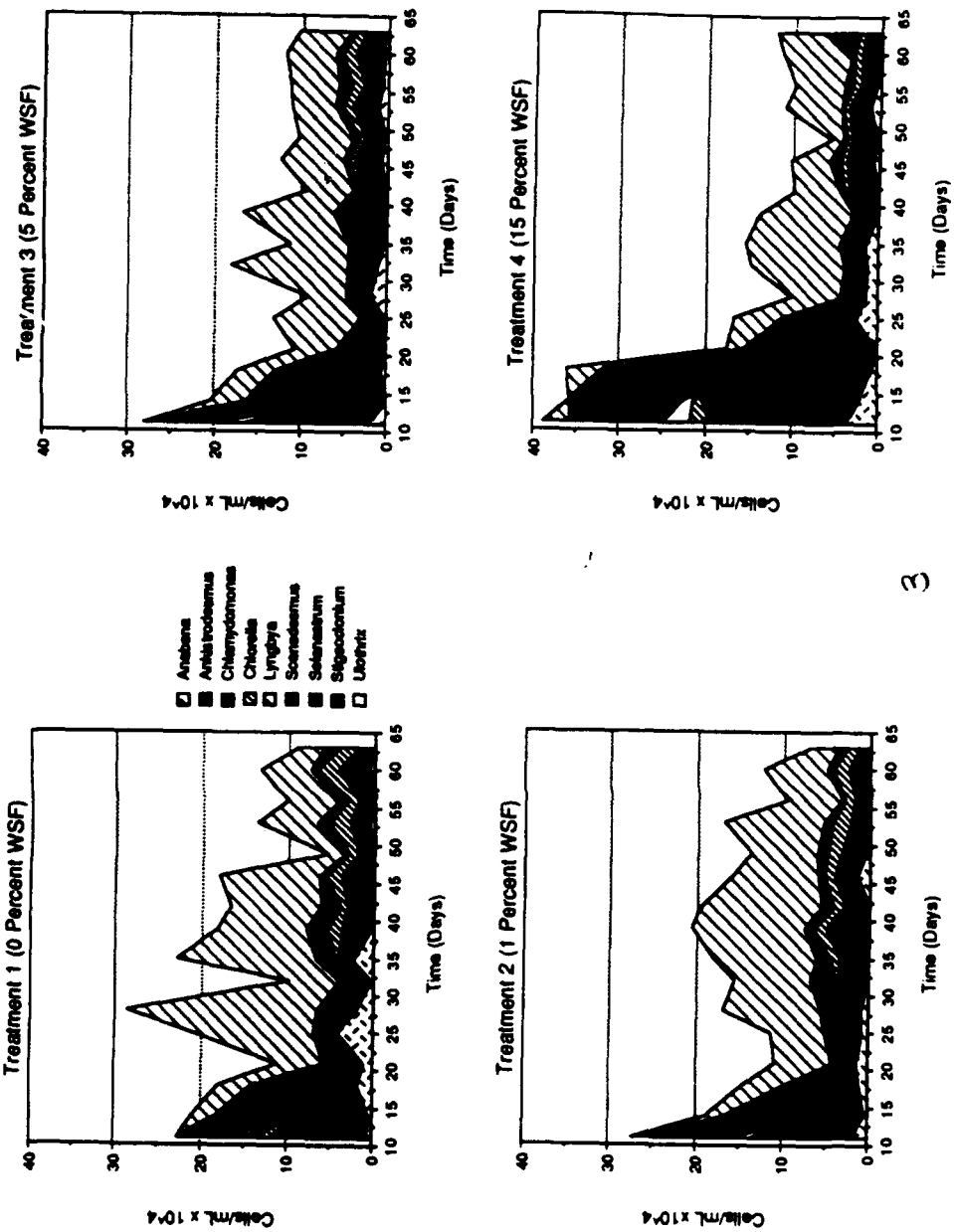


Fig. 3. Patterns in algal communities. The largest increase in algal population density occurred in treatment 4. The peak density is approximately twice that of the control replicates at day 21. After the initial bloom in treatment 4 no particular dose-related pattern is discernible.

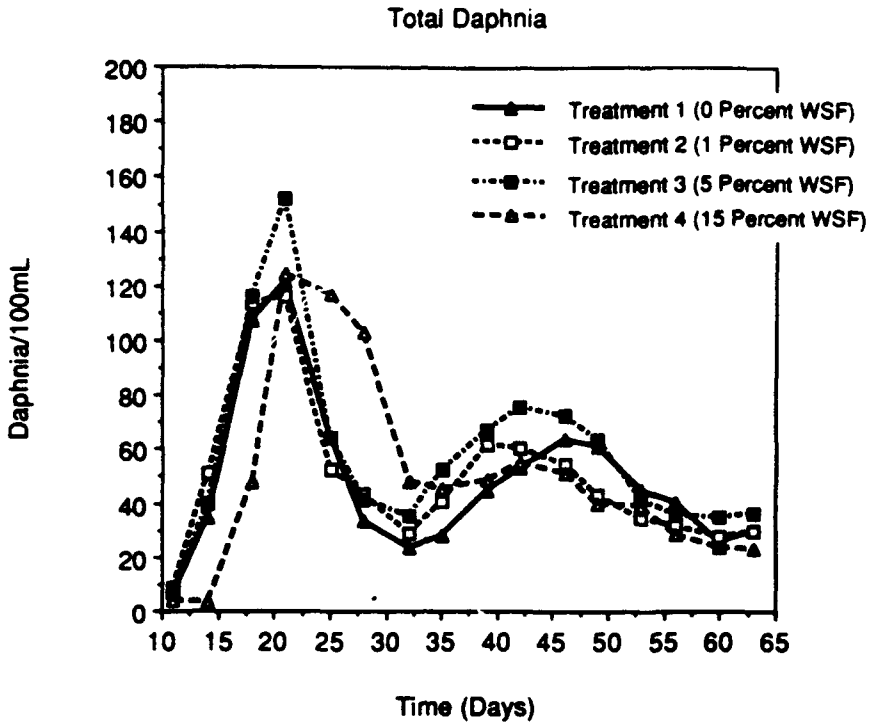


Fig. 4. Daphnid population dynamics. Each of the treatment groups exhibited similar dynamics. None of the groups were statistically different from the control groups using conventional analysis of variance and IND approaches. Minor perturbations in the timing of the peaks may have occurred, but by day 49 the means of each group are very similar.

pH and photosynthesis/respiration ratio.

Treatment 4 pH did exhibit a statistically significant difference from the other treatments during the period of the algal bloom during the first ten days after dosing (Fig. 7). On day 49 a deviation from the control in a dose response manner was detected. However with the multiple comparisons being made it is difficult to attribute such an event to the treatment. At the end of the experiment all of the groups resembled reference treatment.

The photosynthesis/respiration ratio did not exhibit statistically significant differences during the course of this experiment.

Multivariate results

The multivariate methods used in the analysis include cosine and vector distances and nonmetric clustering. Cosine distance in a clustering based on the relative cosine from the origin of the multivariate space. The assumption is that similar replicates are close when relative angles are compared. Vector distance assumes that the replicates that form a cluster are near when distances are compared. Nonmetric clustering is a technique

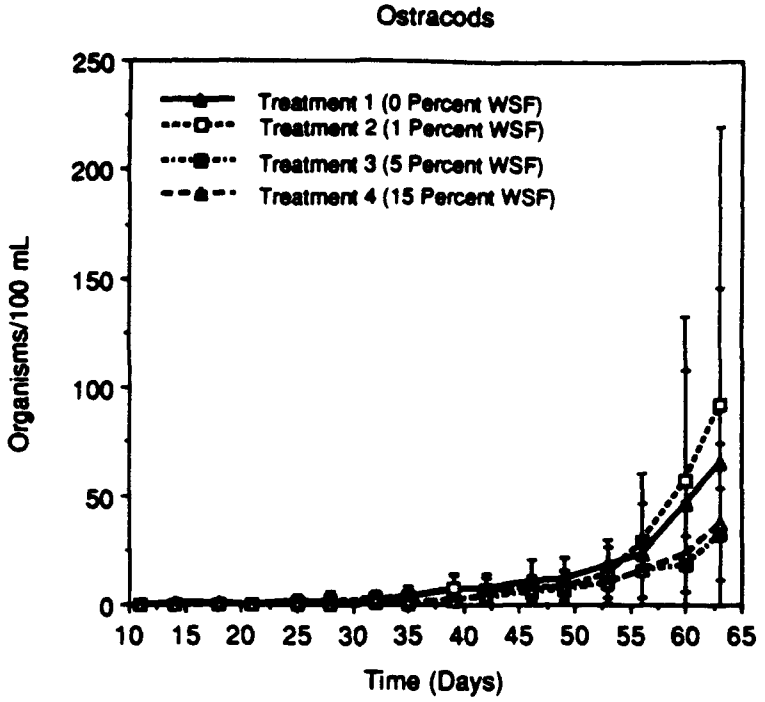


Fig. 5. Ostracod population dynamics. The average population density in the control treatments is approximately twice that of Treatment 4, the highest concentration. In between, the populations densities are ranked in a dose response manner. Although suggestive and not readily apparent in the other biological data, the apparent dose response falls within the IND plot surrounding the control. The bars are standard deviations for the means of each sampling day. An IND is approximately 2.5 times the standard deviation.

where replicates that form clusters have similar characteristics, units of measurement or assumptions as to distribution are not used in this technique.

The significance levels for the three multivariate tests performed for each sampling day are graphed in Fig. 8. All tests agree that a significant difference between treatment groups was observed through day 25. Nonmetric clustering demonstrated fluctuations in this significance from day 25 until 40, and from 40 until the end of the experiment. The cosine vector and Euclidean vector methods were statistically significant until after day 53.

In Fig. 9, the average cosine distances within the reference group and between the reference group and each of the three treatment groups are plotted on a log scale. The initial, strong effect, from day 11 to day 25, is easily seen as a large distance from the reference treatment 1 (no dose) and treatment 4 (highest dose). The period from day 25 to 30 reflects another more subtle oscillation that is statistically significant using cosine vector and Euclidean vector clustering. From day 35 to day 46 the distances from treatment 1 to the other treatments are similar to the within treatment 1 distances and the nonmetric clustering does not detect a significant difference. A third period of

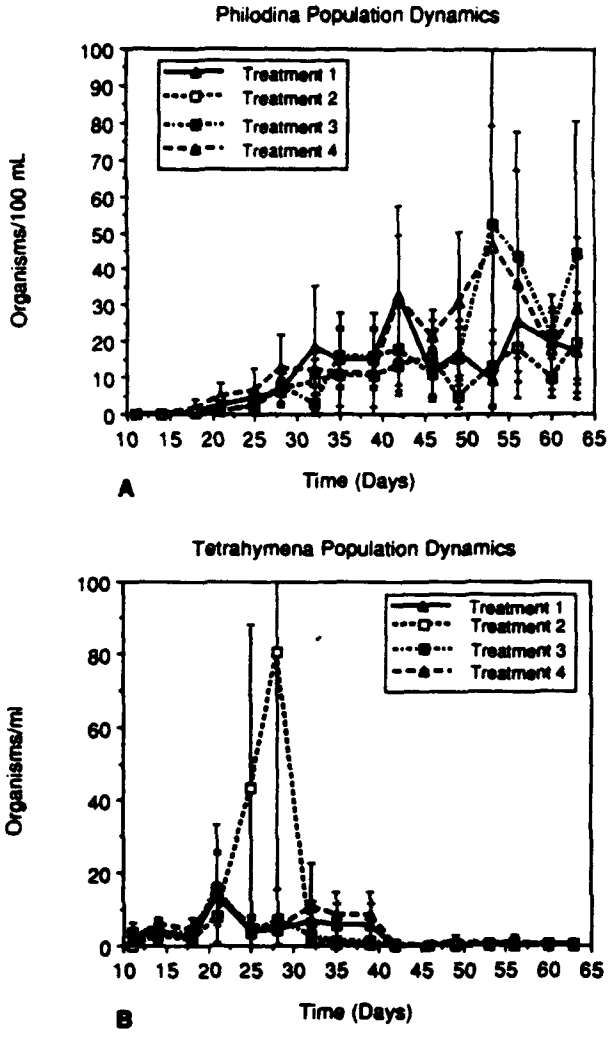


Fig. 6. *Tetrahymena* and *Philodina* population dynamics. The population dynamics of the *Philodina* suggest a treatment effect towards the end of the experiment. As with the ostracods the sampling error is too large to distinguish such an effect using conventional univariate techniques. The bars are standard deviations for the means of each sampling day. An IND is approximately 2.5 times the standard deviation.

separation from the control that is statistically significant using the distance measures, from day 46 to 53, is seen for the JP-4 SAM.

Also of interest are the variables that best described the clusters and the stability of the importance of the variables during the course of the experiment. Table 3 lists the variables determined to be important in defining the clusters of importance for each sampling day as determined by nonmetric clustering. In general, the number of variables

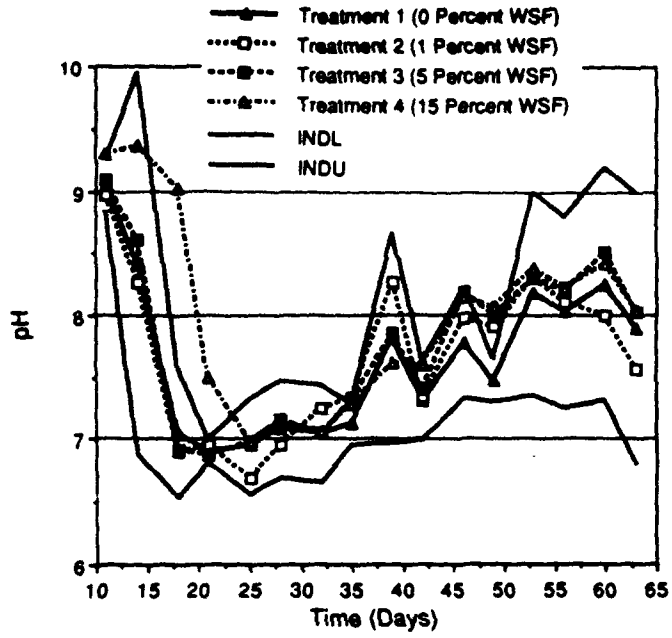


Fig. 7. pH. Treatment 4 pH did exhibit a statistically significant difference from the reference treatment during the period of the algal bloom during the first ten days after dosing (INDL = IND upper limit, INDV = IND upper limit). On day 49 an additional deviation from the control in a dose response manner was detected.

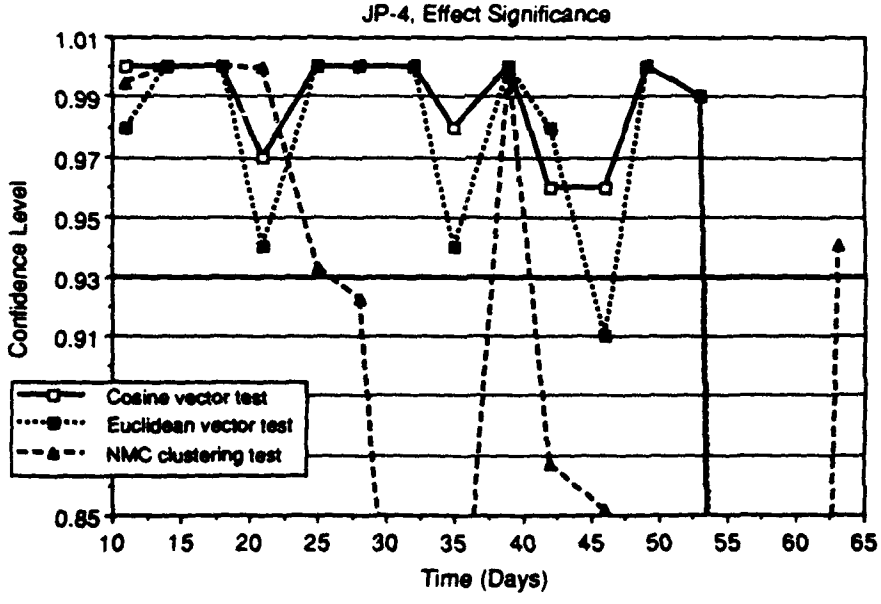


Fig. 8. Significance levels of the three multivariate statistical tests for each sampling day. Note that there are two periods, early and late ones, where the clustering into treatment groups is significant at the 95% confidence level or above.

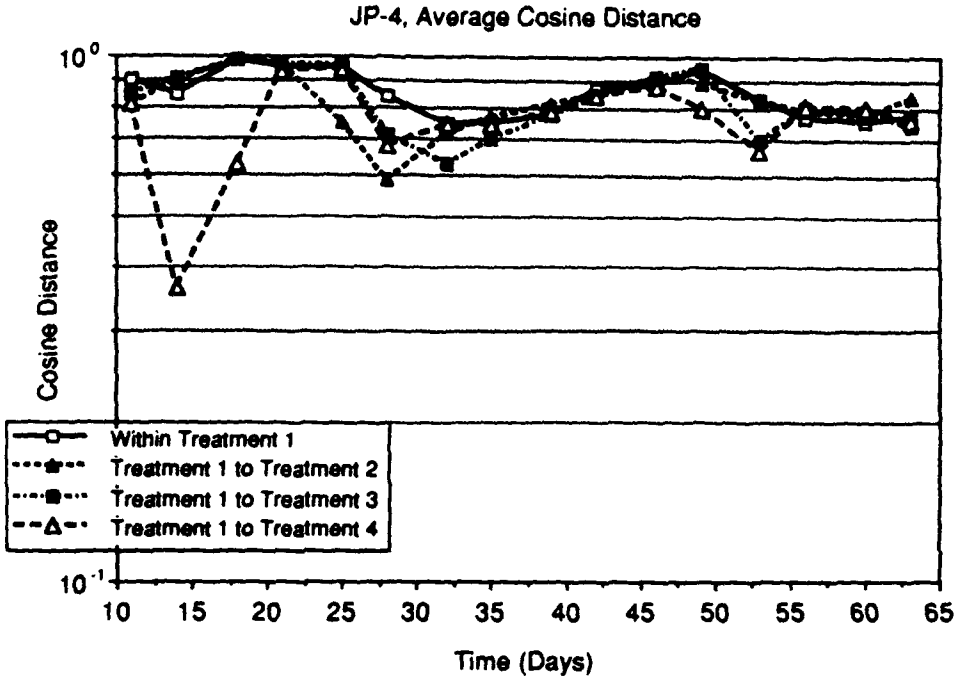


Fig. 9. Cosine distance from the control group to each of the treatments for each sampling day. Note that large differences are apparent early in the SAM. During the middle part of the 63 day experiment the distances between the replicates of Treatment 1, the control group, is as large as the distances to the treatment groups. However, later in the experiment the distances from the dosed microcosms to the control again increase followed by another apparent convergence.

that were important was larger during the start of the test and lower at the end. In addition, a great deal of variability in rankings is apparent during the course of the SAM. The number of sampling dates when a variable was deemed important in cluster formation is listed in Table 4. *Chlorella* and small *Daphnia* were ranked 8 out of the 16 sampling dates with *Ankistrodesmus* ranked 6 out of 16, being ranked in 12 out of the 16 sampling dates. The distribution of ranks was rather even although variables such as *Tetrahymena* and *Ulothrix* did not appear.

The timing of each variable gaining importance in the determination of clusters was also interesting. Ostracods and *Philodina* were important after day 32 of the experiment, as were small *Daphnia*. *Chlorella* was selected as a significant variable throughout the course of the experiment.

Discussion

The examination of individual parameters provided only a limited and somewhat distorted view of the dynamic responses of the SAM system to JP-4. The univariate data did show that there were some significant responses to the toxicant as determined by

Table 3. Important variables as determined by nonmetric clustering ranked according to contribution for each sampling day. Some variables such as *Ankistrodesmus* were important in determining group clusters in the first half of the experiment. Some of the variables such as Ostracod and *Philodina* were more important in the latter stages of the experiment. Note that the order of importance of even the more common contributors often changed from sampling day to sampling day, with no one variable being consistently ranked. *Chlorella* and small *Daphnia* being the closest.

Day	Important Variables in Determining Clusters in Rank Order
11	*Selanastrum*, medium *Daphnia*, *Chlorella*, *Ankistrodesmus*
14	*Selenastrum*, small *Daphnia*, medium *Daphnia-Ankistrodesmus*[a], age *Daphnia-Stigeoclonium*[a]
18	*Scenedesmus*, *Selanstrum*, *Ankistrodesmus*, small *Daphnia*, *Chlorella*, large *Daphnia*
21	*Scenedesmus*, *Ankistrodesmus*, *Chlamydomonas*
25	*Chlorella*, small *Daphnia*
28	*Chlorella*, *Ankistrodesmus-Lyngbya*[a], *Philodina*
32	Ostracod
35	Ostracod, *Philodina*, *Scenedesmus*
39	*Scenedesmus*, small *Daphnia*
42	*Lyngbya*, small *Daphnia*, *Philodina*, *Ankistrodesmus*
46	Medium *Daphnia*
49	*Scenedesmus*, *Chlorella*, *Philodina*
53	*Chlorella*, *Philodina*
56	Medium *Daphnia*-small *Daphnia*[a]
60	Small *Daphnia*, Ostracod, *Lyngbya*
63	*Chlorella*, *small Daphnia*, medium *Daphnia*, *Lyngbya*

[a]Hyphen between variables denotes equal rank.

Table 4. Variable according to success in determining clusters as defined by nonmetric clustering. Variables such as *Ankistrodesmus* and the *Daphnia* classes were important in the course of this study. However, reliance on any particular organism or a small combination would have poorly described the dynamics of the system

Variable	Ranked
Chlorella	8
Small *Daphnia*	8
Ankistrodesmus	6
Scenedesmus	5
Philodina	5
Medium *Daphnia*	4
Lyngbya	4
Large *Daphnia*	3
Ostracod	3
Selenastrum	3

the chemistry. Biological data, taken individually, did not demonstrate a coherent and unified picture of the response of the biota to JP-4. The biological responses that were most evident were of only dramatic impacts, such as the increase in the algal populations due to the inhibitory effect of the JP-4 upon the grazer populations. Axiomatically, an inhibition of the predominant grazer in the early stages of the microcosm, the *Daphnia*, is going to result in an algal bloom. These types of responses do not provide a depth of understanding of the function and structure of the artificial ecosystem. In contrast to the biological data, pH did demonstrate some statistically significant differences using the IND methodology that hinted at an early major impact in treatment 4 and a later divergence. It is likely that pH is measuring an alteration in the metabolism of the system and therefore a change in the functionality, but without structural differences it is difficult to attribute the functional differences to structural alterations.

The multivariate analyses of the structural data revealed patterns not observed using the univariate analysis of the biotic data. Three oscillations from the non dosed treatment 1 could be observed that were statistically significant. Two of these oscillations correspond well to the divergences seen in the pH analysis. However in the divergences seen between days 25–30 and 50–55 (Fig. 9), suggestions of a dose-response can be seen that are not apparent in the pH data. It is important that these oscillations were observed after the demise of the original WSF mixture, no doubt lost to volatilization or biotransformation and degradation by the biota.

Comparison of jet fuel microcosms

A similar set of results have been obtained for a related toxicant, Jet-A (Landis *et al.* 1993). In a virtually identical experiment, univariate methods were able to demonstrate alteration in the grazer (daphnid)-algal dynamics and in two functional measures, pH and P/R ratio. Subsequent departures of the dosed treatments from the non dosed treatments were not observed using the biotic measures. However, the functional measures, pH and P/R, both demonstrated an additional divergence for one sampling date in the latter half of the microcosm experiment. However, the univariate analysis does not corroborate these results and they may have been dismissed as chance occurrences without the multivariate analyses.

The multivariate analyses depicted at least two statistically significant oscillations using all three measurement techniques. As with the Jet-A, the original WSF mixture had rapidly decreased in concentration during the first few weeks after dosing.

A detailed comparison of the dynamics of the two SAM experiments is currently underway to compare similarities and differences in the multivariate space of the impacts of the two mixtures. However, changes in the structural composition of the systems did occur repeatedly during the course of the experiments even in these relatively simple systems. These oscillations point to effects not readily observed or predicted by single species systems. The repeated divergence of the dosed systems from the reference systems can be accounted for in two ways:

(1) It may reflect the functioning of the community in terms of parameters not directly sampled by the SAM protocol.

(2) It may be a persistent fluctuation in the community structure initiated by the initial stress, but is only periodically visible, as if it were an incompletely dampened nonlinear oscillation in the systems' inherent dynamics.

Examination of individual parameters provides only a limited, and somewhat distorted view of the SAM microcosm response to the WSF of each fuel. The univariate data analysis did indeed show that there were some significant responses to the toxicant by individual taxa and chemistry; however, the responses were scattered over time, and did not present a logical, coherent pattern. Furthermore, the individual responses detected were typified by wild swings in the population density of a taxon over time.

If you kill or restrict the reproduction of most of the *Daphnia*, the next microcosm response is probably an algal bloom. This result could easily have been predicted by the short term toxicity tests and was expected. However, recent modelling efforts by Taub *et al.* (submitted) suggest that the dynamics of these interactions and the resulting magnitudes of the algal blooms are highly dependent upon the timing of the toxic insult. Measuring these types of gross responses to the toxicant do not provide much more insight into impact of the toxicant in the ecosystem than do the short-term single-species tests. The absolute magnitude of the disturbance and the period of recovery can be obtained from the microcosm experiment, in the sense of a classical predator prey interaction. However, the multivariate analysis reveals a more interesting dynamic.

The multivariate patterns suggest a much more complex pattern of multiple divergences and convergences in the similarities between treatment groups. Much as an ecosystem could be expected to display the rise and fall of species assemblages, the SAM microcosms appear to indicate that the first divergence is only the beginning of a series of responses.

Using nonmetric clustering, we can list the variables that were the most important for separating the treatment group clusters for each day that measurements were collected (Table 3). The list of variables suggests that the first divergence, which occurred from about day 11 through day 25, results from predator/prey interactions between primary producers (algae) and first order consumers (*Daphnia*). This divergence should be characterized by the following properties:

(1) The divergence will be fast, because the algae and *Daphnia* populations are introduced into the microcosm after being cultured in optimal laboratory conditions and then placed into cultures with high available nutrient concentrations. Predation, or the lack of predation, or other limiting factors will cause rapid changes in the algal and herbivore populations.

(2) The divergence will be short-lived, because the populations are unstable in the nutrient rich early successional microcosm. There will be a tendency for the microcosms to drift away from the early 'treatment' effect into a more typical community based on both algae and detritus as the food source for the secondary consumers. Initially, this drift may mask treatment effects and be interpreted as recovery of the system.

The first divergence is the only type of response that is normally searched for in microcosm tests using conventional statistics. This response is typical of many reported SAM experiments (Taub *et al.* 1988, Taub 1988, Haley *et al.* 1988, Landis *et al.* 1989).

The second and third divergences occurred from between days 25–30 and 50–55. During this time, *Daphnia* and some of the algal taxa were often still important in the cluster development; however, other secondary consumers (Ostracods and *Philodina*) entered the list. The second divergence may represent the long-term effects of the initial toxicant on a more successionally mature community that is fuelled by both algal

productivity and detritus. If so, the resulting divergences should have the following characteristics:

(1) It should be strongly influenced by detritus quality. Detritus is conditioned by bacteria and fungi, which are highly sensitive to toxicants but are unmeasured in the microcosm. Also, detritus that has passed through the gut of a consumer (e.g., consumed algae) is different from detritus that originates directly from dead algae (unconsumed). Therefore, the quality of the detritus may be highly affected by the treatment, but none of the factors influencing the effects will be measured directly.

(2) Secondary consumers of detritus and bacteria are no less affected by the quality of their food source than algal consumers, so the treatment-related alterations of the quality of detritus and bacteria will cause differences in the secondary consumer populations.

Therefore, the series of divergences following the initial algal-daphnid interaction may still represent a direct response to the initial treatment effects, but because it occurs late in the microcosm experiment, it is easily misinterpreted as noisy or the effects of a degradation product. An inclusion of measures of detritus quality and microbial metabolism may answer these questions and such studies are currently being incorporated into our series of microcosm experiments.

Invoking unseen properties of an ecosystem or other mechanistic explanations may not be needed to explain the occurrence of oscillations and divergences from a non-dosed reference system. An alternative and complimentary explanation is available that perhaps describes the dynamics of multispecies systems at a more fundamental level.

Ecosystem dynamics and the illusion of recovery

The return of a system to its pre-existing state, structurally, metabolically and dynamically, is a classical definition of recovery. There are many biotic and abiotic factors that govern the composition of an ecosystem after a stress event; substrate type, distance from colonizing sources, genetic variability of the resident population are but a few. Since each of the initial conditions are likely to be different from those that lead to the original system, it is unlikely that the subsequent system will be identical. Similarity, however, does not mean the same. In fact, similarity at the structural level may lead to an illusion of recovery.

First, the apparent recovery or movement of the dosed systems towards the reference or treatment 1 case may be an artifact of our measurement systems that allow the n-dimensional data to be represented in a two dimensional system. In an n-dimensional sense, the systems may be moving in opposite directions and simply pass by similar coordinates during certain time intervals. Positions may be similar, but the n-dimensional vectors describing the movements of the systems can be very different.

The apparent recoveries and divergences may also be artifacts of our attempt to choose the best means of collapsing and representing n-dimensional data into a two or three dimensional representation. In order to represent such data, it is necessary to project n-dimensional data into three or fewer dimensions. As information is lost when the shadow of a cube is projected upon a two dimensional screen, a similar loss of information can occur in our attempt to represent n-dimensional data. The possible illusion of recovery based on this type of projection is diagramatically represented in Fig. 10. In Fig. 10a the dosed and the reference systems appear to converge, i.e. recovery

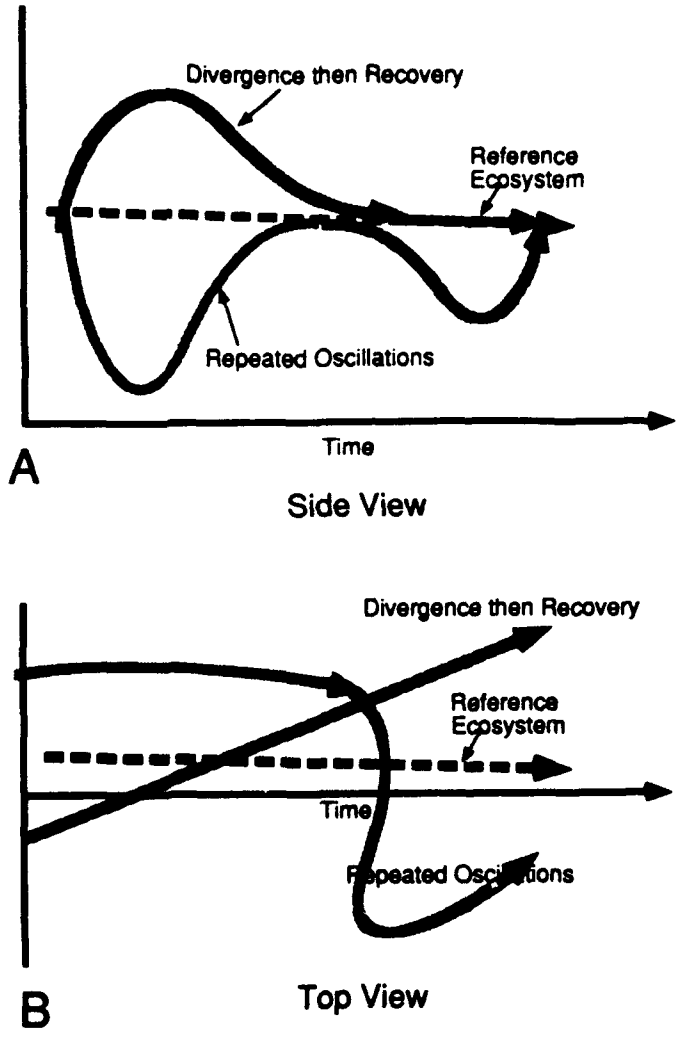


Fig. 10. Diagrammatic representation of ecosystem movements in ecosystem space. In a, the dosed and the reference systems appear to converge, i. e. recovery has occurred. However, this may be an illusion of the variables chosen to describe the system. Fig. 10b is the same system but viewed from the 'top'. When a new point of view is taken, divergence of the systems occurs throughout the observed time period.

has occurred. However, this may be an illusion created by the perspective chosen to describe and measure the system. Figure 10b is the same system but viewed from the 'top'. When a new point of view is taken, divergence of the systems occurs throughout the observed time period. As the various groups separate, the divergence may be seen as a separate event. In fact, this separation is a continuation of the dynamics initiated earlier upon one aspect of the community. Eventually, the illusion of recovery may simply be the divergence of the replicates within each treatment group becoming large

enough, with enough inherent variation, so that even the multivariate analysis can not distinguish treatment group similarities. Not every divergence from the control treatment may have a causal effect related to it in time: differentiating these events from those due to degradation products or other perturbations is challenging.

Complexity in nonlinear ecological systems

Not only may system recovery often be an illusion but strong theoretical reasons indicate that recovery to a reference system may be impossible or at least unlikely. In fact, systems that differ only marginally in their initial conditions and at levels probably impossible to measure are likely to diverge in unpredictable manners. May and Oster (1978) in a particularly seminal paper investigated the likelihood that many of the dynamics seen in ecosystems, generally attributed to chance or stochastic events, are in fact deterministic. In fact simple deterministic models of population can give rise to complicated behaviours. Using equations resembling those used in population biology, bifurcations occur resulting with several distinct outcomes. Eventually, given the proper parameters, the system appears chaotic in nature although the underlying mechanisms are completely deterministic. Obviously, biological systems have limits, extinction being perhaps the most obvious and best recorded. Another ramification is that the noise in ecosystems and in sampling may not be the result of a stochastic process but the result of underlying deterministic, chaotic relationships.

These principles also apply to spatial distributions of populations as recently reported by Hassell *et al.* (1991). In a study using host-parasite interactions as the model, a variety of spatial patterns were developed using the Nicholson-Bailey model. Host-parasite interactions demonstrated patterns ranging through static 'crystal lattice' patterns, spiral waves, chaotic variation or extinction with the appropriate variation of only three parameters within the same set of equations. The deterministic patterns could be extremely complex and not distinguishable from stochastic environmental changes.

Given the perhaps chaotic nature of populations it may not be possible to predict accurately species presence, population interactions, or structural and functional attributes. Katz *et al.* (1987) examined the spatial and temporal variability in zooplankton data from a series of five lakes in North America. Much of the analysis was based on limnological data collected by Brige and Juday from 1925 to 1942. Copepods and cladocera, except *Bosmina*, exhibited larger variability between lakes than between years in the same lake. Some taxa showed consistent patterns among the study lakes. They concluded that the controlling factors for these taxa operated uniformly in each of the study sites. However, in regards to the depth of maximal abundance for calanoid copepods and *Bosmina*, the data obtained from one lake had little predictive power for application to other lakes. Part of this uncertainty was attributed to the intrinsic rate of increase of the invertebrates with variability increasing with a corresponding increase in r_{max}. A high r_{max} should enable the populations to accurately track changes in the environment. Katz *et al.* suggest that these taxa be used to track changes in the environment. Unfortunately, in the context of environmental toxicology, the inability to use one lake to predict the non-dosed population dynamics of these organisms in another, reduces the sensitivity of methods that use comparisons of two systems as measures of anthropogenic impacts.

A better strategy may be to let the data and a clustering protocol identify the important parameters in determining the dynamics of and impacts to ecological systems.

This approach has been recently suggested independently by Dickson *et al.* (1992) and Matthews and Matthews (1991). This approach is in direct contrast to the more usual means of assessing anthropogenic impacts. One classical approach is to use the presence or absence of so called indicator species. This assumes that the tolerance to a variety of toxicants is known and that chaotic or stochastic influences are minimized. A second approach is to use hypothesis testing to differentiate metrics from the systems in question. This second approach assumes that the investigators know *a priori* the important parameters. Given that the important parameters in differentiating non-dosed from dosed systems change from sampling period to sampling period, this assumption can not be made. Classification approaches such as nonmetric clustering or the canonical correlation methodology developed by Dickson *et al.* eliminate these assumptions.

Implications for monitoring and risk assessment

The results presented in this report combined with the others cited above and the implications of chaotic dynamics suggest that reliance upon any one variable or an index of variables is an operational convenience that may provide a misleading representation of pollutant effects and the associated risks. The use of indices such as diversity and the Index of Biological Integrity have the effect of collapsing the dimensions of the descriptive hypervolume in a relatively arbitrary fashion. Indices, since they are composited variables, are not true endpoints. The collapse of the dimensions that are composited tends to eliminate crucial information, such as the inherent variability, and its importance in describing these variables. The mere presence or absence and the frequency of these events can be analysed using techniques such as nonmetric clustering that preserve the nature of the dataset. A useful function was certainly served by the application of indices, but the new methods of data compilation, analysis and representation derived from the Artificial Intelligence tradition can now replace these approaches and illuminate the underlying structure and dynamic nature of ecological systems. In the next 12 months RISC (reduced instruction set computer) based personal computers will make these approaches widely available and rapidly run at the desktop.

The implications are important. Currently, only small sections of ecosystems are monitored or a heavy reliance is placed upon so-called indicator species. Our data suggest that this is dangerous, potentially producing misleading interpretations and resulting in costly error in management and regulatory judgments. Much larger toxicological test systems are currently analysed using conventional statistical methods on the limit of acceptable statistical power. Interpretation of the results has proven to be difficult.

The dynamics observed in our experiments and in the research discussed above should make obvious that a metaphor such as ecosystem health is inappropriate and misleading. In a recent critical evaluation, Suter (1993) dismissed ecosystem health as a misrepresentation of ecological science. Ecosystems are not organisms with the patterns of homeostasis determined by a central genetic core. Since ecosystems are not organismal in nature, health is a property that can not describe the state of such a system. The urge to represent such a state as health has lead to the compilation of variables with different metrics, characteristics and relationships. Suter suggests a better alternative would be to evaluate the array of ecosystem processes of interest, with an underlying understanding that the fundamental nature of these systems is quite different from those of organisms.

One of the ongoing debates in environmental toxicology has been the suitability of

the extrapolation and realism of the various multispecies toxicity tests that have been developed over the last 15 years. One of the major criticisms of small scale systems is that the low diversity of the system is not representative of natural systems in dynamic complexity (Sugiura 1992). Given the above discussion and the conclusions derived from it much of this debate may have been misdirected. The small scale systems used in our study have been demonstrated to express complex dynamics. Kersting and Van Wungaarden (1992) found that even the three compartment microecosystem, as developed by Kersting (1984, 1985, 1988), expresses indirect effects as measured by pH changes after dosing with chloropyrifos. Since even full scale systems can not serve as reliable predictors of the dynamics of other full scale systems, it is impossible to suggest that any artificially created system can provide a generic representation of any full scale system. Debate should probably revert to more productive areas such as improvements in culture, sampling and measurement techniques or other characteristics of these systems. A more worthwhile goal is probably the understanding of the scaling factors, in a full n-dimensional representation, that should enable the accurate representation of specific ecosystem characteristics. Certain aspects of a community may be included in one system to answer specific questions that in another system would be entirely inappropriate. If questions as to detritus quality are important then the system should include that particular component. In other words, the system should attempt to answer the particular scientific question.

Several questions are now the goals of future research. The dynamics of the loss of jet fuels from the SAM systems is currently being investigated in greater depth. Additional data should indicate the persistence of the constituents and help aid in the determination of initial toxicity, including further information from literature searches or using quantitative structure activity relationship models. Additional testing of related materials is being conducted. Finally, questions as to the effects of size and community structure abound. The SAM system is relatively simple. Data sets incorporating more diverse species assemblages and of varying sizes are being investigated for comparison.

Conclusion

Effects are seen in the microcosm after the degradation of the toxicant to very low levels in an oscillating pattern of divergence from the non-dosed treatment, apparent recovery, which is then followed by another divergence.

Multivariate analysis is crucial in observing effects with typically noisy datasets and points to the dynamic nature of the variables important in distinguishing the four treatment groups. Univariate methods would have discounted the contributions of variables such as *Philodina* and Ostracod, since the dosed treatments could not be demonstrated as being statistically different using conventional methods. However, the nonmetric clustering and association analysis demonstrated the importance of these two variables and allows the generation of a new hypothesis, the switch of the system to a detritus base and the resultant differences in system dynamics as indirect effects of the toxicant addition.

Two general hypotheses are proposed to account for the observed dynamics of the system. The oscillations may be the result of structural and functional components not measured, such as detrital processing and quality. The second and not exclusive hypothesis is that the oscillations are due to the inherent nonlinear nature of ecosystems and

may propagate in an inherently unpredictable but not unbounded fashion over time. Nonlinear or chaotic dynamics do not imply random behaviour. In fact, chaotic equations are perfectly deterministic. However, small changes in initial conditions give rise over time to different outcomes. The possibly nonlinear nature of ecosystems does place a time constraint over which, given an initial accuracy of initial conditions, the dynamics of the system can be predicted. Predictions are perhaps better represented as forecasts over specified periods of time.

The implications of these results is that reliance upon indices that condense data or upon indicator species may be misleading in determining effects of stressors upon biological communities. A strategy providing better resolution in determining ecosystem impacts may be the sampling of a broader set of variables, accepting the variability inherent in sampling. Given the difficulty of accurately determining initial conditions, and therefore the dynamics of the system, it may be impracticable or impossible to accurately predict relevant measurements at specific times. If it is inherently impossible to predict the relevant parameters at a specified time, only an examination of a compendium of data from the system is likely to reliably measure effects. A focus on only a few assessments and their corresponding measurement endpoints will probably miss important changes in ecosystem structure and function that create the illusion of sameness, but important differences in the dynamics of structural changes may go undetected.

If multiple undampened oscillations and even chaotic dynamics characterize ecosystems then concepts such as ecosystem health and ecosystem recovery should be eliminated or redefined. Nonlinear or complex systems bordering on the chaotic are unlikely to exhibit characteristics that correspond to health at the organismal level. Similarly, recovery of a system to a preexisting state, both in location and dynamics, may be impossible or highly unlikely.

Appendix A.

Multivariate Techniques-Nonmetric Clustering

In the research described above, three multivariate significance tests were used. Two of them were based on the ratio of multivariate metric distances within treatment groups versus between treatment groups. One of these is calculated using Euclidean distance and the other with cosine of vectors distance (Good 1982; Smith *et al.* 1990). The third test used nonmetric clustering and association analysis (Matthews *et al.* 1990). In the microcosm tests there were four treatment groups with six replicates, giving a total of 24. This example is used to illustrate the applications in the derivations that follow.

Treating a sample on a given day as a vector of values, $\vec{x} = /x_1 \ldots x_n\text{-}$, with one value for each of the measured biotic parameters, allows multivariate distance functions to be computed. Euclidean distance between two sample points $\vec{x}$ and $\vec{y}$ is computed as

$$(1) \qquad \sqrt{\sum_i (x_i - y_i)^2}$$

The cosine of the vector distance between the points $\vec{x}$ and $\vec{y}$ is computed as

$$(2) \qquad 1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum y_i^2}}$$

Subtracting the cosine from one yields a distance measure, rather than a similarity measure, with the measure increasing as the points get farther from each other.

The within-between ratio test used a complete matrix of point-to-point distance (either Euclidean or cosine) values. For each sampling date, one sample point $\bar{x}$ was obtained from each of six replicates in the four treatment groups, giving a 24 × 24 matrix of distances. After the distances were computed, the ratio of the average within group metric (W) to the average between group metric (B) was computed (W/B). If the points in a given treatment group are closer to each other, on average, than they are to points in a different treatment group, then this ratio will be small. The significance of the ratio is estimated with an approximate randomization test. This test is based on the fact that, under the null hypothesis, assignment of points to treatment groups is random, the treatment having no effect. The test, accordingly, randomly assigns each of the replicate points to groups, and recomputes the W/B ratio, a large number of times (500 in our tests). If the null hypothesis is false, this randomly derived ratio will (probably) be larger than the W/B ratio obtained from the actual treatment groups. By taking a large number of random reassignments, a valid estimate of the probability under the null hypothesis is obtained as $(n + 1)/(500 + 1)$ where n is the number of times a ratio less than or equal to the actual ratio was obtained (Noreen 1989).

In the clustering association test, the data are first clustered independently of the treatment group, using nonmetric clustering and the computer program RIFFLE (Matthews and Hearne 1991). Because the RIFFLE analysis is naive to treatment group, the clusters may, or may not correspond to treatment effects. To evaluate whether the clusters were related to treatment groups, whenever the clustering procedure produced four clusters for the sample points, the association between clusters and treatment groups was measured in a 4 × 4 contingency table, each point in treatment group i and cluster j being counted as a point in frequency cell ij. Significance of the association in the table was then measured with Pearson's X^2 test, defined as

$$(3) \qquad \chi^2 = \sum_{ij} \frac{(N_{ij} - n_{ij})^2}{n_{ij}}$$

where N_{ij} is the actual cell count and n_{ij} is the expected cell frequency, obtained from the row and column marginal totals N_{+j} and N_{i+} as

$$(4) \qquad n_{ij} = \frac{N_{+j} N_{i+}}{N}$$

where $N = 24$ is the total cell count, and a standard procedure for computing the significance (probability) of X^2 taken from Press (1990).

Acknowledgements

We would like to thank the support staff of the Institute of Environmental Toxicology and Chemistry, notably L. Keel, R. Sahakian and N. Shough. L. Holmquist and T. Litwiller contributed to the editing and production of this document. This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

References

ASTM D3710 (1988) Standard Test Method for Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography. *1988 Annual Book of ASTM Standards*, Vol. 5.03, pp. 78–88. Philadelphia: American Society for Testing and Materials.

ASTM D2887 (1988) Standard Test Method for Boiling Range Distribution of Petroleum Fractions by Gas Chromatography. *1988 Annual Book of ASTM Standards*, Vol. 5.02, pp. 506–13. Philadelphia: American Society for Testing and Materials.

ASTM E 1218 (1991) Conducting Static 96-h Toxicity Tests with Microalgae. *1991 Annual Book of ASTM Standards*, Vol. 11.04, pp. 845–56. Philadelphia: American Society for Testing and Materials.

ASTM E 1366–91 (1991) Standard Practice for the standardized aquatic microcosm: fresh water. *1991 Annual Book of ASTM Standards*, Vol 11.04, pp. 1017–51. Philadelphia: American Society for Testing and Materials.

Conquest, L.L. and Taub, F.B. (1989) Repeatability and reproducibility of the Standard Aquatic Microcosm: Statistical properties. In Cowgill, U.M. and Williams, L.R. eds. *Aquatic Toxicology and Hazard Assessment, 12th Volume, ASTM STP 1027*, pp. 159–77. Philadelphia, PA: American Society for Testing and Materials.

Crow, M.E. and Taub, F.B. (1979) Designing a microcosm bioassay to detect ecosystem level effects. *Intern. J. Environmental Studies* **13**, 141–7.

Dickson, K.L., Waller, W.T., Kennedy, J.H. and Ammann, L.P. (1992) Assessing the relationship between ambient toxicity and instream biological response. *Env. Tox. Chem.* **11**, 1307–22.

Fisher, L. (1992) *Memorandum: Decisions on the Ecological, Fate and Effects Task Force*. Office of Pesticides and Toxic Substances, US Environmental Protection Agency.

Good, I.J. (1982) An index of separateness of clusters and a permutation test for its significance. *J. Statist. Comp. Simul.* **15**, 81–4.

Haley, M.V., Johnson, D.W. and Landis, W.G. (1988) The aquatic toxicity of brass dust. In Adams, W., Chapman, G. and Landis, W.G., eds. *Aquatic Toxicology and Environmental Fate; 10th Volume ASTM STP -971*, pp. 468–79. Philadelphia: American Society for Testing and Materials.

Hassell, M.P.H., Comins, N. and May, R.M. (1991) Spatial structure and chaos in insect population dynamics. *Nature* **353**, 255–8.

Johnson, A.R. (1988a) Evaluating ecosystem response to toxicant stress: a state space approach. In Adams, W.J., Chapman, G.A. and Landis, W.G., eds, *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP 971*, pp. 275–85. Philadelphia: American Society for Testing and Materials.

Johnson, A.R. (1988b) Diagnostic variables as predictors of ecological risk. *Environmental Management* **12**, 515–23.

Katz, T.K., Frost, T.M. and Magnuson, J.J. (1987) Inferences from spatial and temporal variability in ecosystems: Long-term zooplankton data from lakes. *Amer. Nat.* **129**, 830–46.

Kersting, K. (1984) Development and use of an aquatic micro-ecosystem as a test system for toxic substances. Properties of an aquatic micro-ecosystem IV. *Int. Revue ges. Hydrobiol.* **69**, 567–607.

Kersting, K. (1985) Properties of an aquatic micro-ecosystem V. Ten years of observations of the prototype. *Verh. Internat. Verein. Limnol.* **22**, 3040–5.

Kersting, K. (1988) Normalized ecosystem strain in micro-ecosystems using different sets of state variables. *Verh. Internat. Verein. Limnol.* **23**, 1641–6.

Kersting, K. and van Wungaarden, R. (1992) Effects of Chlorpyifos on a microecosystem. *Env. Tox, Chem.* **11**, 365–72.

Kindig, A.C., Loveday, L.C. and Taub, F.B. (1983) Differental sensitivity of new versus mature

synthetic microcosms to streptomycin sulfate treatment. In Bishop, W.E., Cardwell, R.D. and Heidolph, B.B. eds. *Aquatic Toxicology and Hazard Assessment. 6th Volume, ASTM STP-802*, pp. 192–203. Philadelphia: American Society for Testing and Materials.

Landis, W.G., Savage, R.E. Jr. and Hoskin, F.C.G. (1985) An organofluorophosphate-hydrolyzing activity in *Tetrahymena thermophila*. *J. Protozool.* **32**, pp. 517–9.

Landis, W.G., Haley, D.M., Haley, M.V., Johnson, D.W., Durst, H.D. and Savage, R.E. Jr. (1987) Discovery of multiple organofluorophosphate hydrolyzing activities in the protozoan *Tetrahymena thermophila*. *J. of Appl. Toxicol.* **7**, 35–41.

Landis, W.G., Chester, N.A., Haley, M.V., Johnson, D.W., Muse, W.T. Jr., and Tauber, R.M. (1989) The utility of the standard aquatic microcosm as a standard method for ecotoxicological evaluation. In Suter, G. and Adams, M., eds. *Aquatic Toxicology and Environmental Fate, 11th Volume ASTM STP-1007*, pp. 353–67. Philadelphia: American Society for Testing and Materials.

Landis, W.G. (1991) Distribution and nature of the aquatic organophosphorus acid anhydrases: enzymes for organophosphate detoxification. *Rev. in Aquat. Sci.* **5**, 267–85.

Landis, W.G., Haley, M.V. and Chester, N.A. (1993a) The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In Landis, W.G., Hughes, J. and Lewis, M., eds. *Environmental Toxicology and Risk Assessment, 1st Volume, ASTM STP -1179*, Philadelphia: American Society for Testing and Materials. (in press).

Landis, W.G., Matthews, R.A., Markiewicz, A.J., Shough, N.J. and Matthews, G.B. (1993b) Multivariate analysis of the impacts of turbine fuel using a standard aquatic microscosm toxicity test. *J. Env. Sci.* (In press).

Matthews, G.B., Matthews, R.A. and Landis, W.G. (1990) Applications of nonmetric clustering (NMC) to pattern analysis in aquatic toxicology. *Proceedings of the 11th Annual Society of Environmental Toxicology and Chemistry Meeting*, Arlington, V.A., 1990.

Matthews, G.B. and Hearne, J. (1991) Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 175–84.

Matthews, G.B. and Matthews, R.A. (1991) A model for describing community change. In *Pesticides in Natural Systems: How Can Their Effects Be Monitored? Proceeding of the Conference*, Corvallis, O.R.: Environmental Research Laboratory/ORD, EPA 9109/9–91/011.

Matthews, G.B., Matthews, R.A. and Hachmoller, B. (1991a) Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Can. J. Fish. Aq. Sci.* **48**, 2184–90.

Matthews, R.A., Matthews, G.B. and Ehinger, W.J. (1991b) Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modeling* **53**, 167–87.

May, R.M. and Oster, G.F. (1978) Bifurcations and dynamical complexity in simple ecological models. *Amer. Nat.* **110**, 573–99.

Noreen, E.W. (1989) *Computer Intensive Methods for Testing Hypotheses*. New York, N.Y., Wiley-Interscience.

Press, W.H., Flannery, B.P., Teukolsky, A.A. and Vetterline, W.T. (1990) *Numerical Recipes in C, the Art of Scientific Computing*. New York, NY: Cambridge University Press.

Smith, E.P., Pontasch, K.W. and Cairns, J.Jr., (1990) Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.* **24**, 507–14.

Sugiura, K. (1992) A multispecies laboratory microcosm for screening ecotoxicological impacts of chemicals. *Env. Tox. Chem.* **11**, 1217–26.

Suter, G. (1993) A critique of ecosystem health: Concepts and indices. *Environ Tox. Chem.* (In press).

Taub, F.B. (1969) Gnotobiotic models of freshwater communities. *Verh Internat. Verein. Limnol.* **17**, 485–96.

Taub F.B. (1976) Demonstration of pollution effects in aquatic microcosms. *Intern. J. Environmental Studies* **10**, 23–3.

Taub, F.B. (1988) Standardized aquatic microcosm – development and testing. In A. Boudou and F. Ribeyre, ed., *Aquatic Ecotoxicology* II.

Taub, F.B. (1989) Standardized aquatic microcosms. *Environm. Sci. Technol.* **23**, 1064–6.

Taub, F.B. and Crow, M.E. (1978) Loss of a critical species in a model (laboratory) ecosystem. *Verh. Internat. Verein. Limnol.* **20**, 1270–6.

Taub, F.B., Crow, M.E. and Hartmann, H.J. (1980) Responses of aquatic microcosms to acute mortality. In Giesy, J.P. Jr., *Microcosms in Ecological Research*, Washington, D.C.: Technical Information Center, US Department of Energy, pp. 513–35.

Taub, F.B., Kindig, A.C. and Conquest, L.L. (1987) Interlaboratory testing of a standardized aquatic microcosm. In Adams, pp. 385–405. W.J., Chapman, G.A. and Landis, W.G., eds. *Aquatic Toxicology and Hazard Assessment; 10th Volume, ASTM STP-971*, Philadelphia, P.A.: American Society for Testing and Materials.

Taub, F.B., Kindig, A.C., Conquest, L.L. and Meador, J.P. (1988) Results of the interlaboratory testing of the Standardized Aquatic Microcosm protocol. In Suter, G. and Lewis, M., eds. *Aquatic Toxicology and Hazard Assessment; 11th Volume, ASTM* pp. 368–394. Philadelphia, PA: American Society for Testing and Materials.

Taub, F.B., Rose, K.A., Swartzman, G.L. and Taub, J.H. (submitted) Translating population toxicity to community effects. *Env. Tox. Chem.*

Westendorf, R.G. (1986) Performance aspects of volatile organics analysis by purge and trap capillary column gas chromatography with flame ionization detectors. *Tekmar Technical Papers*. Cincinnati, Ohio: Tekmar Company.

Running head: Uncertainty Propagation in Risk Assessment

Address correspondence to:

Geoffrey Matthews

Computer Science Department, MS 9062

Western Washington University

Bellingham, Washington 98225

Office: (206) 676-3797

Fax: (206) 676-3044

Uncertainty Propagation in Risk Assessment

Geoffrey B. Matthews

Computer Science Department, MS 9062

Western Washington University

Bellingham, Washington 98225

Robin A. Matthews

Huxley College of Environmental Studies, MS 9079

Western Washington University

Bellingham, Washington 98225

Wayne G. Landis

Huxley College of Environmental Studies, MS 9079

Western Washington University

Bellingham, Washington 98225

James W. Hearne

Computer Science Department, MS 9062

Western Washington University

Bellingham, Washington 98225

Footnotes

(none)

Abstract

Risk assessment typically proceeds by successively combining various uncertain inferences into an overall probability. For example, in computing the potential effect on a target species, an extrapolation may have to be made from an acute test on a similar species. A test on white mice, for example, may be pressed into service to estimate effects on deer mice. The expected exposure may be chronic rather than acute, and this will introduce further uncertainty. The test may have been an LC 50 test, while the criteria standards may involve NOELs, which again have to be uncertainly estimated from the LC 50. Typically these uncertainties are combined into a single inferential step, often by assuming worst case in each step, and independence of each uncertainty. This procedure results in a conservative estimate, but rarely an accurate one. Further, it can create an unwarranted variance of several orders of magnitude from the actual test results. This type of inference procedure constitutes a probabilistic reasoning system, for which a number of mathematical formalisms have been developed in the artificial intelligence tradition, such as Dempster-Shafer theory, truth maintenance systems, and nonmonotonic logic. In this paper, we use several cases to illustrate the differences between the conventional approach and a more sophisticated approach that takes into account possible interactions between the various uncertainties in the system. It is generally possible to get much more realistic bounds on the risk assessment by invoking mathematical methods more sensitive to the logic of combined probabilities.

Keywords: uncertainty, risk assessment, probability, artificial intelligence, expert systems

> Life is the art of drawing sufficient conclusions from insufficient premises. —*Samuel Butler*

1 Introduction

Risk assessment involves the combination of a wide variety of more or less uncertain sources of information. Some are known very accurately, such as the gravitational constant or the balances required in redox equations, others are known approximately, such as the LC 50 of copper sulfate for rodents, while others are largely informed conjecture, such as the strength of a public reaction to a 10% increase in the acidity of rain or the stability of an ecosystem. Usually, each of these uncertainties is modelled by a probability distribution over the possible values that each of the variables or parameters of interest can obtain. We discuss here several approaches to uncertain reasoning that come out of the artificial intelligence (AI) tradition, and how use of these techniques might improve the practice of risk assessment.

The variables that go into a risk assessment can be grouped into three major categories:

1. Physical parameters.
2. Decisions.
3. Values.

Physical parameters are things like temperature, pH, number of organisms, and so on. In purely scientific studies, as opposed to policy making studies, physical pa-

rameters are often the only variables that go into the analysis. Decision parameters are items that are under the user's control. The decision to grant permits, for example, can take on such values as: no permits, a few restricted permits, or permits granted to all who apply. The values of the physical variables often feed into the decisions, but generally decisions are made in the hope of maximizing the value parameters. Value parameters are things like jobs, clean air, and healthy wildlife populations.

Establishing reasonable values for these uncertain quantities is a difficult enough task. However, even after the experiments or surveys have been done, the problem remains of *combining* various uncertain quantities, of reasoning from one unsure foundation to another. For example, one may have reasonably accurate information about the relation of a toxin to a particular species, and reasonably accurate information about the structure of the toxin and its toxic relationship to various metabolic pathways, but need to extrapolate this evidence to other species, to an entire ecosystem, or to other toxins. Methodologies such as the QSAR, for example, are attempts to extrapolate from tested species to untested species, *e.g.* rats to *Daphnia*, or from tested compounds to untested compounds, *e.g.* 2.4 dichlorophenol to 2.6 dichlorophenol (Enslein and Craig, 1978; Enslein et al., 1983; Enslein et al., 1988).

Typically, it is assumed that the uncertainties in an analysis are probabilities of one sort or another, and that, accordingly, the only appropriate models for combining them are the laws of probability. However, anylyzing a set of variables (including, perhaps, physical parameters, decisions, and values) with a mathematical, probabilistic model leads quickly to four major problems:

1. A combinatorial explosion of possibilities.

2. A lack of semantic information to guide inferences.

3. Poor methods of dealing with ignorance as well as uncertainty.

4. The need to calculate all values in the model at once, rather than incrementally as evidence is obtained.

Recent AI research has directly addressed these problems. In this paper we briefly consider some of the merits and problems of three AI approaches: localized approaches (which attempt to solve the combinatorial explosion problem), causal nets (which attempt to solve the semantic problem), and Dempster-Shafer calculus (which attempts to solve the ignorance problem). All of them have the benefit of being incremental approaches: as each new piece of information is added to the model, the model incorporates it without large-scale recomputation of all that has gone before.

After a brief introduction to the underlying probabilistic model of uncertainty analysis, we will discuss each of the three AI approaches in turn.

2 Mathematical model

The underlying probabilistic model is well understood in the risk assessment literature (Morgan and Henrion, 1990). If a problem concerns a set of variables, for example $\{A, B, C, D, E\}$, then, for each value that each variable can take on, we need to know the *joint probability* of that combination, $P(a, b, c, d, e)$ (where a is a value A can take on, *etc.*). The immediate problem with this approach is

that it is intractable for even small numbers of variables. If there are, say, only 20 variables in a problem, and each can take on, say, 6 values, then there are $6^{20} = 3,656,158,440,062,976$, over 3 quadrillion, different combinations of these values. Specifying all of these values is plainly unrealistic, but which values are necessary, and which redundant?

If the variables are continuous numbers and can, in effect, take on a infinite number of different values, then the joint probabilities must be specified as continuous multivariate functions of those variables, an even more daunting task. Generally speaking, most practical risk assessment proceeds by making all variables discrete: for example, species may be considered "highly susceptible," "moderately susceptible," or "not susceptible." To keep things simple, we will also, for the most part, assume that variables are categorical, that is, there are only a small number of discrete values they can take on. However, many of the techniques discussed can be generalized to the continuous case.

Characteristically, probabilities are not computed from a full, joint probability distribution, but are dealt with in a probability tree, such as the one in Figure 1. In this figure we have only four variables, and each variable (A, B, C, and D) has two possible values, which we will represent as $+a$, $-a$, *etc.*, and indicate by the upper and lower branches. There are, accordingly, $2^4 = 16$ possibilities, one for each path through the tree from left to right: the ends of the far-right arrows each represent a different possible outcome. The heavy arrows, for example, represent the combination $(+a, -b, -c, +d)$. The numbers on the arrows represent conditional probabilities, based on *all* the choices to the left. For instance, the heavy arrow above C in the figure has the value 0.8, indicating that the conditional

Figure 1 here.

probability of $-c$ given $+a$ and $-b$, is 0.8, written $P(-c|+a,-b) = 0.8$. If all 2^4 probabilities are known in advance (one number attached to each of the ends of the far-right arrows), then these conditional probabilities can be calculated by summing and dividing from right to left. The values at the top right, for example, indicating that $P(+a,+b,+c,+d) = 0.01$ and $P(+a,+b,+c,-d) = 0.004$ together imply that $P(+d|+a,+b,+c) = 0.01/(0.01 + 0.004)$, and so on. Likewise, knowing all of the conditional probabilities will determine the joint probabilities. The heavy arrows, for example, tell us that $P(+a,-b,-c,+d) = (0.3)(0.2)(0.8)(0.1) = 0.0048$.

It is usually much easier for humans to estimate a conditional probability than to estimate a joint probability. For instance, the probability that it rained last night, given that the grass is wet and you heard thunder, could be estimated. But estimating the probability that you will hear thunder tonight and find wet grass in the morning, unconditioned by anything, usually leads to confusion. Human probabilistic judgements are usually conditional, and therefore probability trees such as the one in Figure 1 are usually filled in along the branches, rather than from the right side.

The tree can, of course, be rearranged, putting B before A, etc., and getting a different set of conditional probabilities ($P(+a|-b)$ instead of $P(-b|+a)$, for instance). However, there are still an insuperably large number of conditional probabilities that must be estimated, and the mathematical model itself gives us no help in determining which are relevant and which irrelevant. Further, if there are some probabilities in the tree about which we are largely, or even completely, ignorant, *some* values for them will have to be provided, even if they are completely arbitrary. In situations of complete ignorance, a uniform probability distribution is usually as-

sumed: all outcomes equally likely. Other situations require a "seat-of-the-pants" estimate: for example, we may estimate that 75% of the local population is likely to favor a pesticide regulation, using only the current political climate as guidance. This is not total ignorance, but it is just as arbitrary.

These problems: huge numbers of possibilities, not knowing which of them are relevant, treating ignorance in an *ad hoc* manner, and the basic need to recalculate everything when any one thing changes, lead us into several models of reasoning under uncertainty that stem from the AI tradition. We now turn to a consideration of three of them, and their relative merits in dealing with these problems.

3 Local approaches

Early in the development of expert systems, the combinatorial problems associated with inference under uncertainty were recognized. While it was recognized that, if the presence of a was evidence for b (*e.g.* $P(b|a)$ was high), then even if we know a is true we still cannot conclude anything about b without knowing if a is the *only information relevant* to b. Another factor, such as c, might completely alter our expectations. For example, elevated temperature in an aquatic system generally connotes reduced dissolved oxygen concentrations because of the inverse relationship between oxygen solubility and temperature. However, the elevated temperature may also imply that it is mid-summer. Photosynthetic activity during this time may cause increased dissolved oxygen levels if the values come from the epilimnion of a biologically productive lake (see Figure 2).

Figure 2 here.

Because it was clearly unrealistic for every inference to consult every possibly

relevant fact in the system, an approximate approach was used, which would go ahead and make inferences from a to b, but would attach "certainty factors" to the conclusions. Certainty factors are definitely *not* probabilities: calculating probabilities was deemed too hard and certainty factors were a substitute. An example from the MYCIN system follows (Buchanan and Shortliffe, 1984). MYCIN was an early expert system constructed to perform medical diagnosis: examine symptoms, recommend further tests, and make inferences as to likely causes.

Each inference rule in MYCIN was expressed as an "if-then" statement with a certainty factor attached, such as these:

1. If a then c (0.4)
2. If b then c (0.6)
3. If c then d (0.8)

which indicated that, for example, if you were reasonably sure about c, then you would be 80% as sure about d. Various combination rules had to be devised when chains of reasoning were involved. For example, if a and b were both known for certain, the first two rules could be combined under the following formula to get a certainty factor for c:

$$\begin{aligned} CF(c) &= 0.4 + 0.6 - (0.4)(0.6) \\ &= 0.76 \end{aligned}$$

Given this certainty factor for c, the third rule above could be used to give a certainty factor for d:

$$CF(d) = (0.76)(0.8)$$

$$= 0.61$$

The MYCIN certainty factors take on both positive and negative values, allowing evidence to be either for or against a conclusion.

Such localized rules essentially solved the combinatorial explosion problem by ignoring it. Their use resulted in practical, working systems that solved large problems in the real world (Buchanan and Shortliffe, 1984). However, they had to be used with great care, because, strictly speaking, their inferences were invalid. Consider, for example, what would happen with these rules if different *types* of reasoning are mixed. Some inferences are from cause to effect; for example, if you open the floodgates, you can safely infer that the water downstream will rise. On the other hand, some inferences are from effect to cause; for example, if you find a large fish kill, you can legitimately raise your expectation of toxins in the water. But putting two such inferences together can be disastrous. Consider:

- If the sprinkler was on then the grass is wet (0.9)
- If the grass is wet then it rained (0.8)

Therefore:

- If the sprinkler was on then it rained

  (0.9 * 0.8 = 0.72)

Each of the two original inferences is quite probable; each of their "if" parts lends support to their "then" parts. The combination of the two, however, is ludicrous.

One attempt to incorporate information such as cause-effect relationships into the process of reasoning under uncertainty is provided by causal nets, considered in the next section.

4 Causal nets

Causal nets, also called Bayesian networks or influence diagrams, are an attempt to retain the original probabilistic model, exemplified in Figure 1, but meet head-on the problem of combinatorial explosion by analyzing the *kinds* of links in the diagram, and reducing the number of calculations that have to be done without sacrificing validity of the inferences (Pearl, 1988).

One of the devices brought to bear on this problem is distinguishing cause and effect, as mentioned at the end of the last section. In Figure 3, the inferences from "sprinkler" to "grass" and from "grass" to "rain" are distinguished by being in the opposite causal direction. Inferences from cause to effect are carried by π-messages, while inferences from effect to cause are carried by λ-messages. (Since we normally have conditional probabilities of effects, given causes, π's are associated with probabilities while λ's are associated with likelihoods, hence the names.) Careful handling of λ and π messages at each point avoids the nonsensical inference from "sprinkler" to "rain", but does so in a way that does *not* require every inference to check every other fact in the system before going ahead. In fact, only in certain, restricted classes of systems does any non-local checking have to be done. Causal "loops" are one example, where, for instance, a single cause can have two effects, but each effect can result in the same symptom. In Figure 4, for instance, the observation of increased chlorophyll would naturally lead to an increased probability of algal enhancement, which should strengthen the probability of *both* an oxygen sag (by a π message) and the probability of some form of nutrient enhancement (by a λ message). However, the oxygen sag should not then send a λ message up the

Figure 3 here.

Figure 4 here.

fish kill → nutrient ladder, because this would increase the probability of nutrients *twice* on the *same* piece of evidence.

Such loops raise problems for the causal net model, and there are a number of approaches to dealing with them; but these problems are minor compared to a straightforward mathematical model which would require *all* factors be reconsidered in *all* inferences.

A number of other advantages to the causal net model come about as well. The importance of *qualitative* uncertainties is obvious. The EPA Framework for Ecological Assessment, for example, asserts that,

> ...often the relationship [between measurement and assessment endpoints] can be described only qualitatively. Because of the lack of standard methods for many of these analyses, professional judgment is an essential component of the evaluation (U. S. Environmental Protection Agency, 1992, p. 23)

However, a causal net model offers a standard, formal, and qualitative treatment of independence. In the mathematical model, for example, independence of events is defined quantitatively, based on the probability distributions: a is said to be independent of b, given c, if and only if

$$P(a|b,c) = P(a|c)$$

Clearly, to establish this in general, one has to go back to the joint probabilities and calculate things numerically. Humans, however, can often judge whether two things are independent, without having the slightest idea of the numeric probabilities involved. Consider, for instance, a watershed study and the question of whether or

not rainfall is independent of soil type. Normally we could easily judge that these two factors are independent. However, to verify this mathematically, the joint probabilities for each plot of land, for each amount of rain, and for each soil type, would all have to be calculated or estimated. This is clearly a large task, and also plainly a waste of time given that we can judge their independence qualitatively without any of the numbers.

Causal nets, on the other hand, by distinguishing π (cause to effect) and λ (effect to cause) inferences, can give deep qualitative insight into this kind of independence. For example, height and reading ability in humans are highly correlated. However, if you know a subject's age (presumably the root cause of the correlation between height and reading ability), then height and reading ability become independent. On the other hand, earthquakes and burglaries are largely independent, but both can cause your car-alarm to go off. Hearing your car alarm simultaneously raises the probability of both a burglary and an earthquake, but also renders them dependent—hearing about an earthquake on your radio will decrease your expectation of a burglar at your car. Rainfall and soil type, for another example, are only *conditionally* independent. If it is learned that a hill slope failure occurred, then rainfall and soil type are no longer independent: a very stable soil type would increase the probability of heavy rain before the failure. Causal nets, in conjunction with algorithmic inference engines, can automate such complex qualitative reasoning. The automation of such inferences becomes critical as the systems dealt with become more complicated, and dozens or hundreds of intertwined causes and effects begin to interact.

An extension of the causal net model to continuous-valued numeric variables is

straightforward (Pearl, 1988, pp. 344-356), and only requires that some tractable model of the uncertainties be used. The usual assumptions about uncertainties, such as uncorrelated, normal distributions, and linear interactions between variables, suffice.

5 Dempster-Shafer theory

Causal nets are an improved reasoning tool for dealing with probabilities such as those found in the standard model (Figure 1). However, even with the improvements found in a causal net approach, at times the probabilities in the standard model remain intractable. Dempster-Shafer theory was designed to overcome some of these problems, by approaching probabilities in an entirely different light (Shafer, 1976; Gordon and Shortliffe, 1984). To understand this approach, consider a standard model with just two variables, a and b. In the standard model, probabilities must be assigned to all possible outcomes, namely, $(+a, +b)$, $(+a, -b)$, $(-a, +b)$, and $(-a, -b)$. Even in a situation of total ignorance, *some* probabilities (such as 0.25 to each) would have to be assigned to these. In the Dempster-Shafer model, *sets* of possible outcomes are considered. Probabilities are defined over these sets, denoting the hypothesis, in each case, that one or another of the possible outcomes in the set will be the true one. In our two variable example, for instance, the sets might consist of such things as $\{(+a, +b), (-a, -b)\}$, denoting the hypothesis that either *both* a and b will be the case, or neither will, or $\{(+a, -b), (-a, +b)\}$, denoting the hypothesis that if either a or b happens, the other won't.

The logic of this approach thus contrasts with the standard model. Rather

than making joint probabilities easier to deal with by breaking them down into conditional probabilities, joint probabilities are simplified by lumping them together. The intuition is that many working hypotheses in science are of this nature: a disease symptom, for example, may indicate one of several diseases and eliminate others. The presence of such a symptom, then, is evidence for an hypothesis that is essentially a disjunction: it's probably either A or B or C, where each of the hypotheses (A, B, and C) is itself a *complete* specification of the system.

This approach has the advantage of immediately simplifying most problems. In dealing with a complex ecological system, for instance, a natural approach does not usually involve hypotheses governing all possible states of all variables in all combinations. Rather, a few models are conjectured that have consequences for *all* of the variables. For example, a eutrophic lake would characteristically imply high temperature, low dissolved oxygen, and a deep depth. An oligotrophic lake, on the other hand, would imply high temperature, high dissolved oxygen, and either deep or shallow depth. More finely divided scenarios would be devised, of course, to fit the level of assessment desired.

Further, the calculation of probabilities over these sets is freed from some of the problems that plague causal nets and other "Bayesian" approaches. The selection of prior probabilities, for example, is eliminated. Rather than, say, assigning a uniform probability to all possible outcomes in the case of complete ignorance, the Dempster-Shafer theorist simply assigns probability one to the set of all possible outcomes (a set usually denoted by Θ, and called the *frame of discernment*), and zero to any subset. To make sure these probabilities of *sets* of hypotheses are not confused with probabilities of hypotheses, we use m instead of P, and say $m(\Theta) = 1.0$. In

a Bayesian approach, by contrast, the initial state of ignorance might be modelled using a uniform distribution: for example, if there were n possible outcomes, each one would be assigned a probability of $1/n$.

For a simple example of subsequent calculations and the incremental propagation of uncertain information in the Dempster-Shafer model, consider a simple situation in which there are only three possible outcomes, A, B, and C. All possible subsets of these outcomes are illustrated in Figure 5 (except the empty set, which, by assumption, will never have a probability greater than 0). The frame of discernment $\Theta = \{A, B, C\}$ is at the top, and the subset relation is indicated by an arrow. Initially,

Figure 5 here.

$$\begin{aligned} m(\Theta) &= 1.0 \\ m(\{A, B\}) &= m(\{A, C\}) = \ldots = 0.0 \end{aligned}$$

(A Bayesian approach, on the other hand, would have $P(A) = P(B) = P(C) = 1/3$.) Now suppose that information is gained suggesting, at a level of 0.6, that either B or C is correct. We update as:

$$\begin{aligned} m(\Theta) &= 0.4 \\ m(\{B, C\}) &= 0.6 \\ m(\{A, B\}) &= m(\{A, C\}) = \ldots = 0.0 \end{aligned}$$

Notice that the remainder ($0.4 = 1.0 - 0.6$) is *not* assigned to $\{A\}$, the *complement* of $\{B, C\}$, but remains with the completely neutral hypothesis set, $\{A, B, C\}$. This accords well with intuitions: evidence in favor of $\{B, C\}$ should not *increase* the probability of $\{A\}$ from 0 to 0.4.

Combining further evidence with this m function proceeds as follows. Let us call the above function m_1, and suppose we gain evidence in favor of $\{A, B\}$, with strength 0.5. This would give us a new function, m_2, with

$$\begin{aligned} m_2(\Theta) &= 0.5 \\ m_2(\{A, B\}) &= 0.5 \\ m_2(\{B, C\}) &= m_2(\{A, C\}) = \ldots = 0.0 \end{aligned}$$

In this case, we would expect B to be supported at some level greater than zero, since it was supported by both pieces of evidence, and this is the case. The combined measure function, m_3, obtained from m_1 and m_2, is defined as follows, for any set Z:

$$m_3(Z) = \sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y)$$

Accordingly,

$$\begin{aligned} m_3(\{B\}) &= m_1(\{B, C\}) \cdot m_2(\{A, B\}) \\ &= (0.6)(0.5) \\ &= 0.3 \\ m_3(\{A, B\}) &= m_1(\{A, B, C\}) \cdot m_2(\{A, B\}) \\ &= (0.4)(0.5) \\ &= 0.2 \\ m_3(\{B, C\}) &= m_1(\{B, C\}) \cdot m_2(\{A, B, C\}) \\ &= (0.6)(0.5) \\ &= 0.3 \end{aligned}$$

$$\begin{aligned} m_3(\{A,B,C\}) &= m_1(\{A,B,C\}) \cdot m_2(\{A,B,C\}) \\ &= (0.4)(0.5) \\ &= 0.2 \end{aligned}$$

and all other m_3 values are zero. Notice that the sum of all m_3 values remains one, as a probability distribution should. Occasionally, when evidence supports mutually incompatable hypotheses, the sum drops below one. For example, if one experiment supported A as the only explanation, and another experiment supported only B, then the empty set, $\emptyset = \{A\} \cap \{B\}$, representing "no possible explanation of the evidence," would get some amount of support. In this case, Dempster-Shafer theory specifies that the probabilities of the nonempty sets are simply scaled up so that the total sum remains one. Thus, the full equation for m_3, given m_1 and m_2, is:

$$m_3(Z) = \frac{\sum_{X \cap Y = Z} m_1(X) \cdot m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) \cdot m_2(Y)}$$

This equation can be applied in an incremental fashion as each piece of information is acquired, or each decision contemplated.

These calculations may appear confusing and involved, and their justification involves deep results in model theory and logic (Shafer, 1976), but they are nonetheless intuitively satisfying and they can be fully automated. The important fact to notice about them is that practitioners, in dealing with uncertain evidence, need only specify which *sets of hypotheses* the evidence supports. The precise impact of a piece of evidence on any one variable, physical parameter, decision, or value, need not be estimated. Combinations of particular variables can be combined into scenarios, and the probabilities of each scenario dealt with directly. This can result in considerable conceptual clarity in dealing with complex situations. The usual

requirements of expert solicitation, that he or she imagine wildly unlikely combinations of events, and then estimate probabilities for other variables conditioned on them, are absent from the Dempster-Shafer methodology. Only likely scenarios, combinations of variable values, need be considered.

6 Conclusion

The logic of combined probabilities, studied extensively in the artificial intelligence tradition, is amenable to a large number of approaches. The mathematical foundations of probability are usually based on building up definitions and theorems based on complete knowledge of a joint probability distribution. However, the higher-level reasoning often pursued by humans in their assessment of uncertainty and risk often has little or no basis in numerical combinations of a huge number of probability estimates. Nevertheless, current practice in risk assessment often assumes that such rock-bottom numbers must be obtained or estimated, by some means, before uncertain inference can proceed.

We have outlined three recent approaches to uncertain inference that stem from the artificial intelligence tradition. Localizing the inferences allows us to forget about many of the numbers involved, but at the expense of making quite unreliable inferences at times. Causal nets reduce some of the complexity of the problem, can support automated qualitative reasoning about uncertainty, and are faithful to the cause/effect distinction which permeates uncertain reasoning. Dempster-Shafer theory allows uncertain reasoning to proceed on a different level, on the level of sets of likely scenarios rather than sets of variables and their values, and as a result

greatly reduces the effort in translating human intuition into an automated system, and has a much more intuitively satisfying treatment of ignorance.

The ability to automate each of these approaches, to embody their inference structure into a computer program, has the potential for even greater rewards. A long tradition of machine learning has found that often a computer-generated analysis can be superior to human intuition. A strong example is provided by Michalski's expert system (Michalski and Chilausky, 1980). Michalski and his colleagues went through a long consultation phase with a human expert in soybean pathology in an effort to build an expert system capable of diagnosing soybean diseases. Michalski then used a machine learning system to build a second expert system solely from *data* concerning soybean diseases and their symptoms; in other words, he used another AI program, a *learning* program, to extract the rules used by the second expert system. Both expert systems were then tested on new cases. The set of rules produced by the human pathologist correctly identified only 83% of the new diseases, while the set of rules produced by the computer program correctly identified 99.5% of the new cases. "...plant pathologists are now using the machine-induced rules for their routine diagnoses" (Firebaugh, 1988).

A recent study of the future of computer science and engineering (CS&E) by a committee of the National Research Council concluded that recent advances in CS&E were not readily available to many other disciplines, and called on CS&E to increase its interactions with other disciplines. Among the top priorities for the future of CS&E they listed:

- Increase its contact and intellectual interchange with other disciplines ...
- Increase the number of applications of computing and the quality of existing applications in areas of economic, commercial, and social significance ...
- Increase traffic in CS&E-related knowledge and problems among academia, industry, and society at large, and enhance the cross-fertilization of ideas in CS&E between theoretical underpinnings and experimental experience

(Committee to Assess the Scope and Direction of Computer Science and Technology, NRC, 1992, p. 34)

This paper is an attempt to initiate a dialogue between CS&E professionals versed in many techniques of automated reasoning under uncertainty and the practitioners of risk assessment nationwide. Each of the approaches sketched here has great potential in risk assessment, particularly in automated software tools which may soon form a critical part of the risk analyst's repertoire.

7 Acknowledgements

This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

References

Buchanan, B. G. and Shortliffe, E. H. (1984). *Rule-Based Expert Systems, The*

MYCIN Experiments of the Stanford Heuristic Programming Project. Addison-Wesley. Reading. Massachusetts.

Committee to Assess the Scope and Direction of Computer Science and Technology. NRC (1992). Computing the future. *Communications of the ACM.* 35(11):30-40.

Enslein. K., Blake. B. W., Tuzzeo. T. M., Borgsdtedt. H. H., Hart. J. B., and Salem. H. (1988). Estimation of rabbit eye irritation scores by structure-activity equations. *In Vitro Toxicology.* 2(1):1-14.

Enslein. K. and Craig. P. N. (1978). A toxicity estimation model. *Journal of Environmental Pathology and Toxicology,* 2:115-121.

Enslein. K., Lander. T. R., Tomb. M. E., and Landis. W. G. (1983). Mutagenicity (ames): A structure-activity model. *Teratogenesis. Carcinogenesis. and Mutagenesis.* 3:503-513.

Firebaugh. M. W. (1988). *Artificial Intelligence: A Knowledge-Based Approach.* Boyd & Fraser Publishing Company. Boston. MA.

Gordon. J. and Shortliffe. E. H. (1984). The Dempster-Shafer theory of inference. In Buchanan. B. G. and Shortliffe. E. H., editors. *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project.* pages 272-292. Addison-Wesley. Reading. Massachusetts.

Michalski. R. S. and Chilausky. R. L. (1980). Learning by being told and learning from examples: an experimental comparison of the two methods of knowledge

acquisition in the context of developing an expert system for soybean diseases. *Policy Analysis and Information Systems*. 4.

Morgan, M. G. and Henrion, M. (1990). *Uncertainty, A Guide to Dealing with Uncertinty in Quantitative Risk and Policy Analysis*. Cambridge University Press, Cambridge.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.

U. S. Environmental Protection Agency (1992). *Framework for Ecological Risk Assessment*. EPA/630/R-92/001.

Legends for Figures

Figure 1. Basic probability model. Each path from left to right represents a combination of the variables A, B, C. and D. Conditional probabilities lie along arrows, joint probabilities are found at the extreme right hand side.

Figure 2. A case in which one cause (high temperature) can lead to different effects in different circumstances. The conditional probability alone of low dissolved oxygen, given high temperature, does not allow an inference from high temperature to low dissolved oxygen.

Figure 3. Bayesian inference takes account of cause and effect by distinguishing inferences based on causes (π inferences) from inferences based on effects (λ inferences).

Figure 4. A causal loop that must be handled carefully in Bayesian inference, even if π anf λ inferences are distinguished.

Figure 5. Dempster-Shafer theory calculates probability over sets of hypotheses, not single variable values. This illustration shows all possible subsets of three hypotheses.

Figures

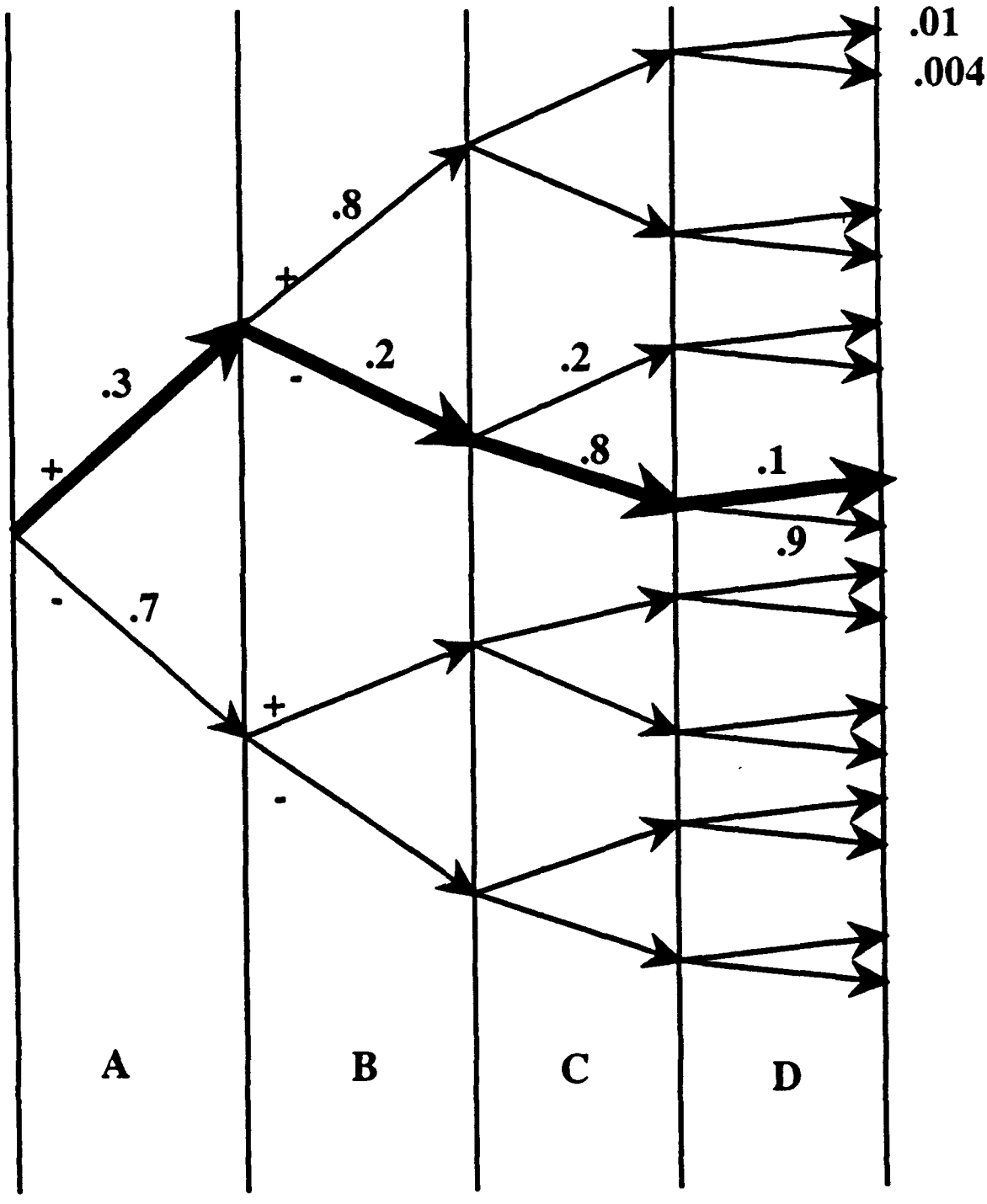


Figure 1

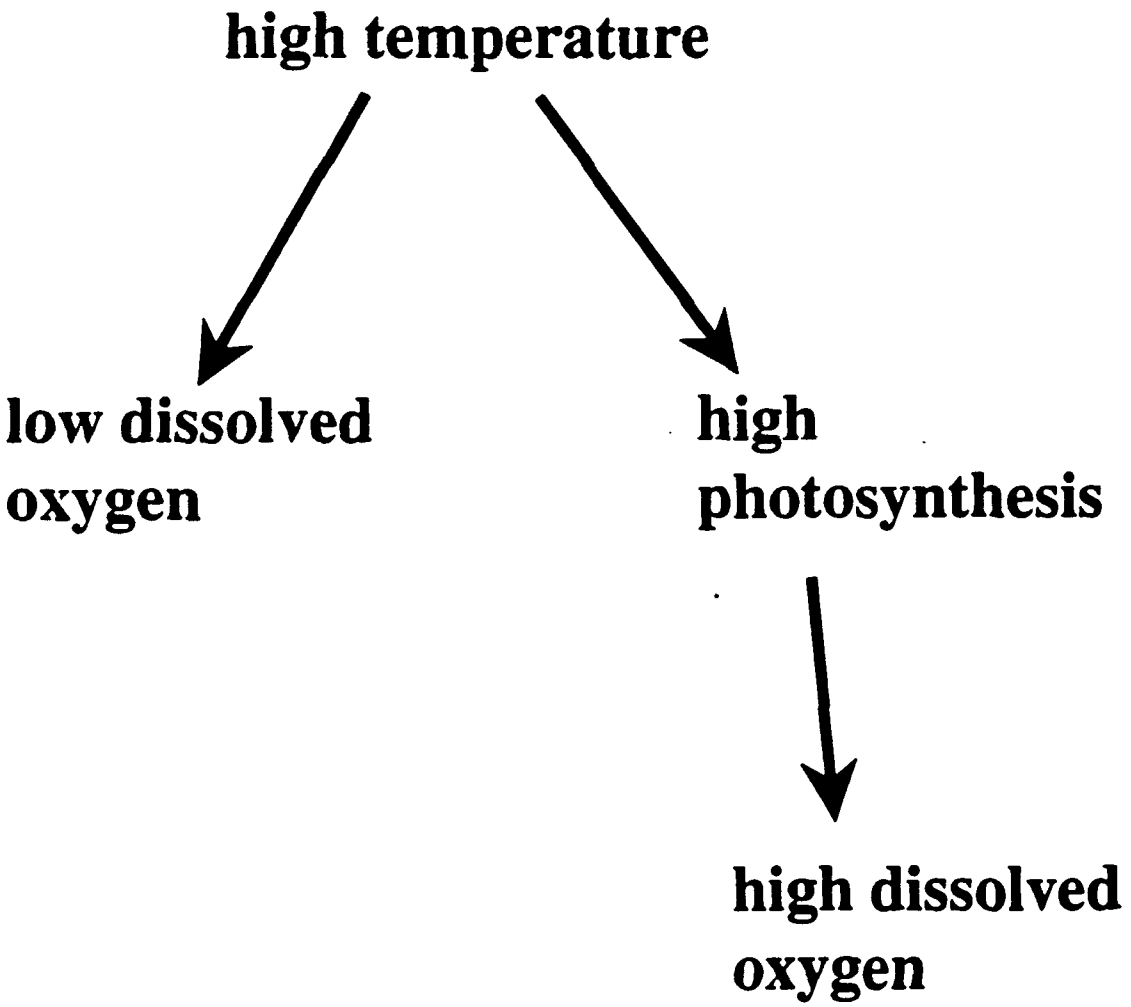

high temperature
low dissolved oxygen
high photosynthesis
high dissolved oxygen

Figure 2

Bayesian Networks

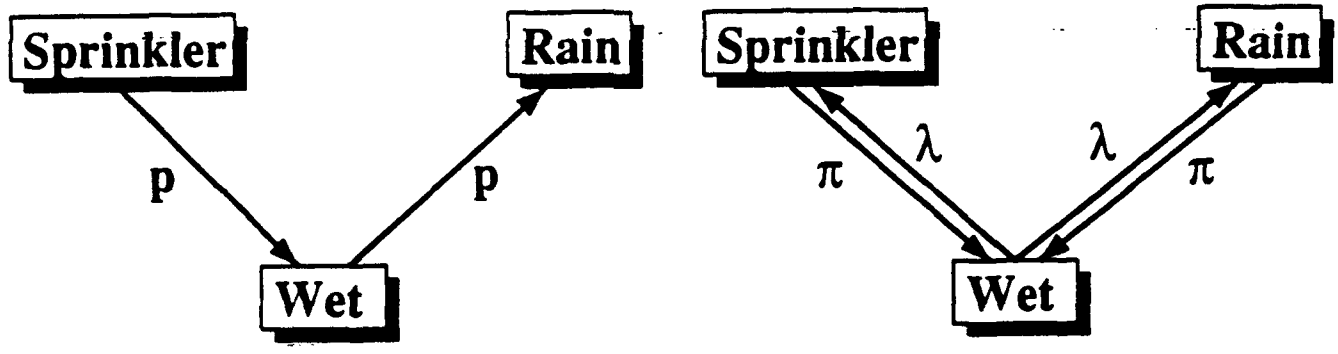


Figure 3

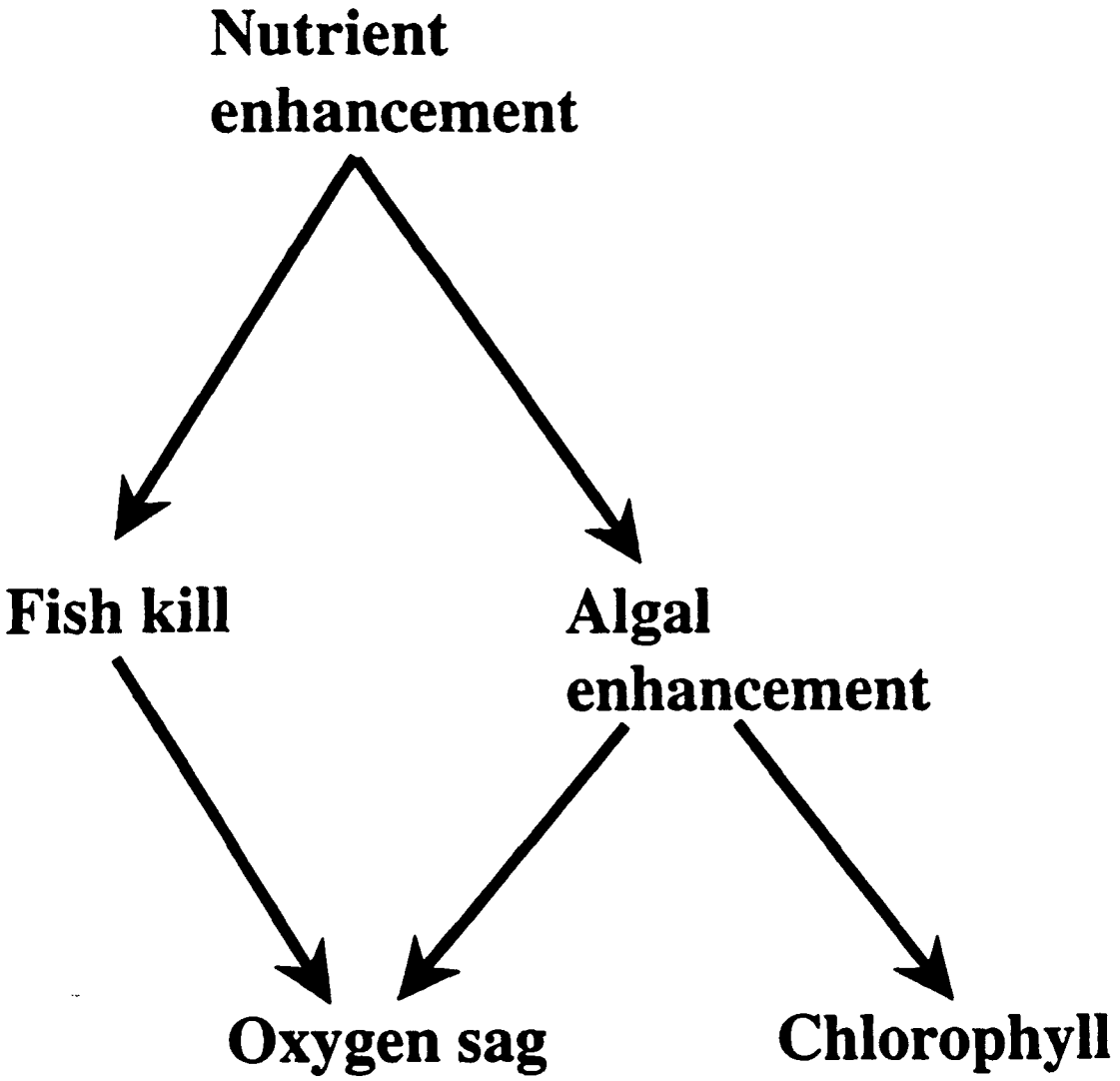

Nutrient enhancement
Fish kill
Algal enhancement
Oxygen sag
Chlorophyll

Figure 4

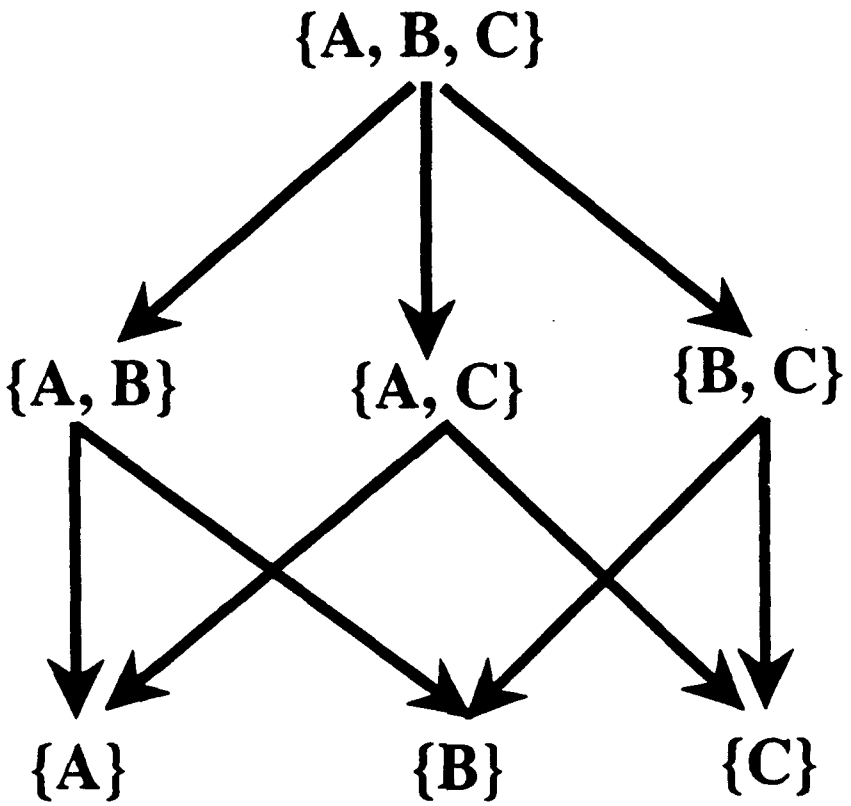


Figure 5

1)Draft Report, Please do not cite or quote

Wayne G. Landis[1], Robin A. Matthews[1], April J. Markiewicz[1] and Geoffery B. Matthews[2].

Non-linear Oscillations Detected By Multivariate Analysis in Microcosm Toxicity Tests with Complex Toxicants: Implications for Biomonitoring and Risk Assessment.

REFERENCE: Landis, W. G., Matthews, R. A., Markiewicz, A. J. and Matthews, G. B. "Non-linear Oscillations Detected By Multivariate Analysis in Microcosm Toxicity Tests with Complex Toxicants: Implications for Biomonitoring and Risk Assessment," Environmental Toxicology and Risk Assessment-Third Volume, ASTM 1218, Jane S. Hughes, Gregory R. Biddinger, and Eugene Mones, Eds., American Society for Testing and Materials, Philadelphia, 1994.

A common assumption in environmental toxicology is that after the initial stress, ecosystems recover to resemble the control state. This assumption may be based more on our inability to observe an ecosystem with sufficient resolution to detect differences, than reality. This study compares the dynamics of the effects of the water soluble fraction (WSF) of both Jet-A and JP-4 using the Standard Aquatic Microcosm (SAM) using several types of multivariate analysis.

Two SAM experiments have been completed using concentrations of 0.0, 1, 5 and 15 percent WSF. The effects of the WSF on the microcosm communities were subtle. Among the more interesting effects were the shifts in time of population peaks and some other variables compared to reference microcosms. In both experiments, multivariate analysis was able to differentiate oscillations that separate the treatments from the reference group, followed by what would normally appear as recovery, followed by another separation into treatment groups as distinct from the reference treatment. These patterns generally were not detected by conventional analysis.

Two sets of related explanations exist for the observed phenomenon. First, the addition of the toxicant initiates an alteration in the community so that the quality of the food resources for the later successional stages is significantly different from the control. This difference in resource quality and quantity leads to the repeated and replicated oscillations. The second explanation is that the oscillations are the result of the intrinsic chaotic behavior of population interactions, of which the alteration of detrital quality is but one of many. The initial impact of the toxicant re-set the dosed communities into different regions of the n-dimensional space where recovery may be an illusion due to the incidental overlap of the oscillation trajectories occurring along a few axes. Some of the implications of non-linear or chaotic dynamics upon the prediction of ecological risk are discussed.

Key Words: Standardized Aquatic Microcosm, jet fuel, non-linear dynamics, nonmetric clustering and association analysis, risk assessment

[1]Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies and
[2]Computer Science Department, Western Washington University, Bellingham, WA 98225.

INTRODUCTION

Over the last 15 years a variety of multispecies toxicity tests have been developed with the hope that in doing so, the increased complexity of the test would result in a more realistic comparison to community-level responses to the toxicant. However, the addition of more than one species, and the generally longer time periods associated with these multispecies tests, also result in much more complex data sets. Distinguishing toxicant effects from other community-level changes has become one of the most critical obstacles to the interpretation of multispecies data sets.

Multispecies toxicity tests are usually referred to as microcosms or mesocosms, although a clear definition of the size or complexity to distinguish these terms has not been put forth. In the Standardized Aquatic Microcosm (SAM) developed by Taub and colleagues (Taub 1969, 1976, 1988, 1989, Taub and Crow 1978, Crow and Taub 1979, Taub et al. 1980, 1987, 1988, Kindig et al. 1983, Conquest and Taub 1989) the physical, chemical, and biological components are defined as to species, media and substrate. The SAM system has undergone round robin testing (Conquest and Taub 1989) and has been used with a variety of toxicants and degradative organisms (Landis et al. 1989, 1993).

One of the major difficulties in the evaluation of multispecies toxicity tests has been the difficulty in the analysis of the large data set on a level consistent with the goals of the toxicity test. Typically, the goals of the multispecies toxicity test are twofold:

- to detect changes in the population dynamics of the individual taxa that would not be apparent in single species tests; and,

- to detect community-level differences that are correlated with treatment groups thereby representing a deviation from the control group.

A number of methods have been developed in an attempt to satisfy the goals of multispecies toxicity testing. Analysis of variance (ANOVA) is the classical method to examine single variable differences from the control group. However, because multispecies toxicity tests generally run for weeks or even months, there are problems with using conventional ANOVA. These include the increasing likelihood of introducing a Type II error (accepting a false null-hypothesis), temporal dependence of the variables, and the difficulty of graphically representing the data set. Conquest and Taub (1989) developed a method to overcome some of the problems by using intervals of non-significant difference (IND). This method corrects for the likelihood of Type II errors and produces intervals that are easily graphed, facilitating further analysis. The method is routinely used to examine data from SAM toxicity tests, and it is applicable to other multivariate toxicity tests. The major drawback of the IND is the limitation of examining one variable at a time over the course of the experiment. While this method addresses the first goal in multispecies toxicity testing, listed above, it ignores the second. In many instances, community-level responses are not as straightforward as the classical predator/prey or nutrient limitation dynamics, that are usually selected as examples of single-species responses representing complex interactions.

Multivariate methods have proved promising as a method of incorporating all of the dimensions of an ecosystem. One of the first methods used in toxicity testing was the calculation of ecosystem strain developed by Kersting (1984, 1985, 1988) for a three compartment

microcosm. This method has the advantage of using all of the measured parameters of an ecosystem to look for treatment-related differences. At about the same time, Johnson (1988a, 1988b) developed a multivariate algorithm using the *n*-dimensional coordinates of a multivariate data set and the distances between these coordinates as a measure of divergence between treatment groups. Both of these methods have the advantage of examining the ecosystem as a whole rather than by single variables, and can track such processes as succession, recovery and the deviation of a system due to an anthropogenic input.

However, a major disadvantage of both these methods, and of many conventional multivariate methods, is that all of the data are often incorporated without regard to the units of measurement, or to the appropriateness of including all variables in the analysis. Random variables indiscriminately incorporated into the analysis, may contribute so much noise that they overshadow variables that do show treatment-related effects.

Ideally, a multivariate statistical test used for evaluating complex data sets will have the following characteristics:

- It will not combine counts from dissimilar taxa or other variable classifications by means of sums of squares, or other *ad hoc* mathematical techniques.

- It will not require transformations of the data.

- It will work without modification on incomplete data sets.

- It will work without further assumptions on different data types.

- Significance of a variable to the analysis will not be dependent on the absolute size of its count, so that taxa having a small total variance, i.e. rare taxa, can compete in importance with common taxa, and taxa with a large, random variance will not automatically be selected, to the exclusion of others.

- It will provide an integral measure of the quality of the analysis, i.e. whether the data set differs from a random collection of points.

- It will, in some cases, identify a subset of the variables that serve as reliable indicators of the physical and biological environment.

Recently developed for the analysis of ecological data, nonmetric clustering is a multivariate derivative of artificial intelligence research, that satisfies all these criteria and has the potential of circumventing many of the problems of conventional multivariate analysis.

In this paper, we use three multivariate techniques to compare patterns in the data sets from two SAM toxicity tests using turbine fuels. The multivariate techniques include two conventional tests based on the ratio of multivariate metric distances (Euclidean distance and cosine of the vector distance), and one relatively new program, RIFFLE, which employs nonmetric clustering and association analysis (Matthews and Hearne 1991). All three of the multivariate techniques have proven useful in analyzing complex ecological data sets (Matthews et al. 1991a, 1991b). Of the three, only nonmetric clustering meets all of the criteria listed above (Matthews and Matthews 1991).

EXPERIMENTAL METHOD

Reagents

All chemicals used in the culture of the organisms and in the formulation of the microcosm media were reagent grade or as specified by the ASTM method.

Jet-A was provided by Fliteline Services of Bellingham, Washington and was refined by Chevron. The sample was obtained from the sample valve used for quality control. The shipment lot was recorded and is on file. JP-4 was supplied by the U. S. Air Force Toxicology Laboratory at Wright Patterson, AFB, Ohio.

Water Soluble Fractions

The water soluble fraction was prepared in glassware washed in nonphosphate soap, rinsed, then soaked in 2N HCl for at least one hour, rinsed ten times with distilled water, dried and finally autoclaved for 30 minutes. Microcosm medium, T82MV, acted as the diluent for the water fraction of the WSF.

Twenty five mL of fuel is added to the two liter separatory funnel, and is agitated as follows: [1] shake separatory funnel for five minutes, releasing built up pressure as necessary; [2] allow funnel contents to remain undisturbed for 15 minutes; [3] shake contents for five minutes, allow to stand 15 minutes; [4] continue same pattern for a total time of one hour; and finally [5] allow separatory funnel contents to remain undisturbed for eight hours. At the end of this procedure the mixture was allowed to stand overnight. The next day all but 100 mL of T82MV/water soluble fraction of jet fuel mixture from the separatory funnel (leaving the lighter, insoluble fuel mixture in the flask) was drained into a cleaned, sterile 1 liter amber glass bottle and capped with a Teflon-lined screw cap. The WSF was used within 24 hours or stored at 4°C for no longer than 48 hours before use as the toxicant mixture.

Gas Chromatography of WSF

This protocol utilizes a Tekmar LSC 2000 Purge and Trap (P&T) concentrator system in tandem with a Hewlett Packard 5890A Gas Chromatograph with a Flame Ionization Detector (FID) (ASTM D3710, D2887, Westendorf 1986). Instrument blanks and deionized distilled water blanks are used to verify the P&T and GC columns cleanliness prior to analysis of samples. A five mL sample is injected into a five milliliter sparger, purged with pre-purified nitrogen gas for eleven minutes and dry purged for four minutes. Volatile hydrocarbons, purged from the sample and collected on the Tenax/Silica Gel column, are desorbed at 180°C directly onto the gas chromatograph SPB-5, 30m x 0.53 mm ID 1.5µm film fused silica capillary column. The column, at 35°C, is held at that temperature for two minutes, increased to 225°C at 12°C/min and held at that temperature for five minutes. A Spectra-Physics 4290 Integrator records the FID signal output of the volatile hydrocarbons that have been separated and eluted from the column by molecular weight. A comparison is then made of the sample chromatograph to n-paraffin and n-naphtha chromatograph standards for sample concentration determinations.

Identification and Quantification of GC Fractions

Qualitative identification of some components in the WSF were determined using a Simulated Distillation (SIMDIS) Calibration Mixture. The ASTM Method D3710 Qualitative Calibration Mixture is the standard

test method for determining the Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography. This mixture was used as a calibration standard to determine the retention times for each known component in the mixture against which unknown components, in the WSF of the fuel mixture, were compared and identified.

SAM Protocol

The 64-day SAM-protocol previously has been described (ASTM E1366). Briefly, the microcosms were prepared by the introduction of ten algal, four invertebrate, and one bacterial species into 3L of sterile defined medium. Test containers were 4 L glass jars. An artificial sediment consisting of 200 g acid washed silica sand, cellulose and 0.5 g of ground chitin is autoclaved in the experimental jar; immersed in a water bath to a point above the level of the sediment during sterilization to prevent breakage.

Numbers of organisms, dissolved oxygen (DO) and pH were determined twice weekly. Room temperature was 20°C ± 2°. Illumination was 80.0 μEm^{-2} sec^{-1} PhAR with a range of 78.6-80.4 and a 12/12 day/night cycle.

Two major modifications were made to the SAM protocol. The first was the means of toxicant delivery. Test material was added on day 7 by stirring each microcosm, removing 450 mL from each container and then adding appropriate amounts of the WSF to produce concentrations of 0, 1, 5 and 15 percent WSF. After toxicant addition, the final volume was adjusted to 3L. No attempt to filter and retain the organisms withdrawn during the removal of the 450 mL was made prior to toxicant addition. All graphs and statistical analysis start with the next sampling day, day 11. The second modification was the substitution, in the JP-4 experiment, of *Tetrahymena thermophila* BIV for the hypotrichous ciliate. The hypotrichous ciliate was becoming increasingly difficult to culture, very likely due to the age of the clone. The results of the JP-4 study demonstrated the suitability of the Tetrahymena for inclusion in the protocol.

Data Analysis

All data were recorded onto standard computer entry forms and checked for accuracy. Parameters calculated included the concentrations of each of the species, DO, DO gain and loss, net photosynthesis/respiration ratio (P/R), pH, algal species diversity, algal biovolume, and biovolume of available algae. The statistical significance of these parameters, compared to the controls, was also computed for each sampling day using the IND plots developed by Conquest. The net photosynthesis/respiration ratio is not derived using ^{14}C methods but by comparing oxygen concentrations before lights on, at the end of the photosynthetic period just before lights off, and then at the next morning, as specified in the standard protocol. The photosynthesis/respiration ratio was then determined by incorporating these measurements.

The multivariate methods used in the analysis include cosine and vector distances and nonmetric clustering. All of these methods have been previously described (Matthews et al. 1991b, Landis et al. 1993) and are reviewed in this volume. Variables used in the multivariate analysis are presented in Table 1.

RESULTS

Persistence of the fuels. In the case of both WSFs, within three weeks after dosing the original material had been volitilized or degraded. In the case of JP-4, benzene, 2,4 dimethylpentane, ethylbenzene, 2-methylpentane, 2-methylpropane, o-xylene and toluene,

TABLE 1. Biotic parameters used in the multivariate statistical tests. Biotic variables such as diversity, available biovolume, and total algal biovolume are not used since they are derived from and therefore not independent of the variables listed below.

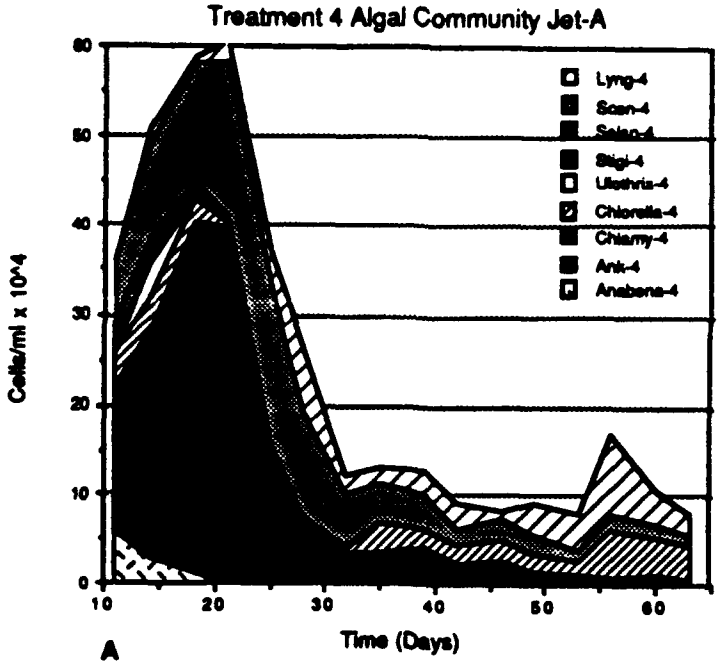
Jet A	JP4
Anabaena	Anabaena
Ankistrodesmus	Ankistrodesmus
Chlamydomonas	Chlamydomonas
Chlorella	Chlorella
Daphnia	Daphnia
Ephipia	Ephipia
Small Daphnia	Small Daphnia
Medium Daphnia	Medium Daphnia
Large Daphnia	Large Daphnia
Hypotricha	Tetrahymena
Lyngbya	Lyngbya
Miscellaneous sp.	Miscellaneous sp.
Ostracod (Cyprinotus)	Ostracod (Cyprinotus)
Philodina (Rotifer)	Philodina (Rotifer)
Scenedesmus	Scenedesmus
Selanastrum	Selanastrum
Stigeoclonium	Stigeoclonium
Ulothrix	Ulothrix

were tracked using GC analysis during the course of the SAM experiment. After week three, only 2-methylpentane and 2-methylpropane are detectable. Since only the 2-methylpropane is present 672 hours after dosing, this material may be the final biodegradative product of the absorbed fraction of the WSF, and is being investigated in more detail.

Comparison of Algal Population Dynamics-Highest Treatment. These area graphs (Figure 1) show the contribution of each algal species to the algal assemblage for the highest treatment concentration for each experiment. In the Jet-A treatment the algal populations were highest, reflecting the increased toxicity of the Jet-A to the daphnid populations. In both experiments however, an algal bloom was observed during the first 30 days of the experiment. At the end of the experiment the numbers and composition of the algal assemblage were similar, although the proportions of the species making up the assemblage had some differences. Chlorella seemed to be a greater constituent of the community in the JP-4 experiment.

Daphnid Population Dynamics. The most direct effect of the jet fuel upon the population dynamics of the daphnid populations was the delay in daphnid reproduction (Fig. 2). Peaks were delayed in the Treatment 4 microcosms in both instances. Daphnids were very important in determining the clusters in the early part of each experiment but not as important later. In both experiments two peaks of daphnid populations are observed. The first reflects the presence of the toxicant, the second occurs similarly in the dosed and not dosed systems. Error bars are not shown for clarity.

Ostracod Population Dynamics. Ostracod populations did not increase until late in each experiment (Fig. 3). In the Jet-A experiment (A), the numbers started an increase between days 40 and 45.



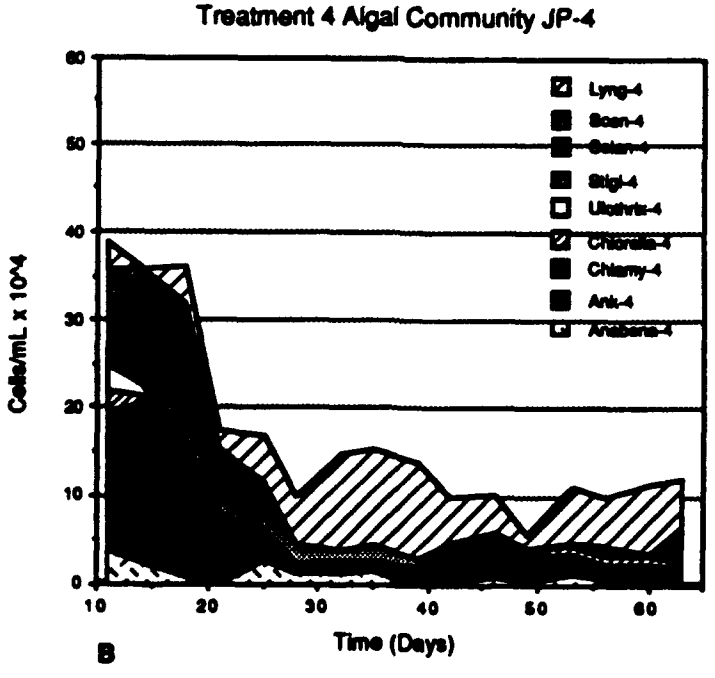
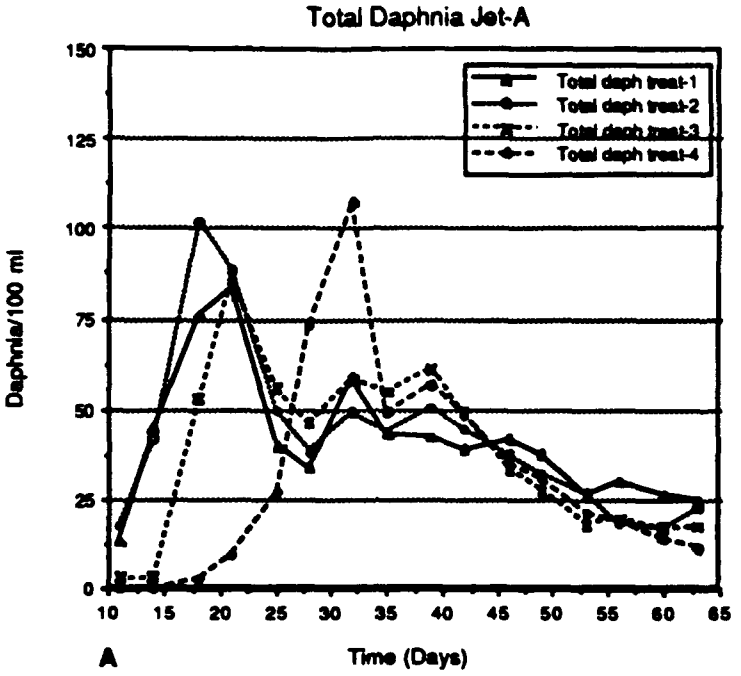


FIG. 1--Comparison of algal population dynamics-highest treatment.



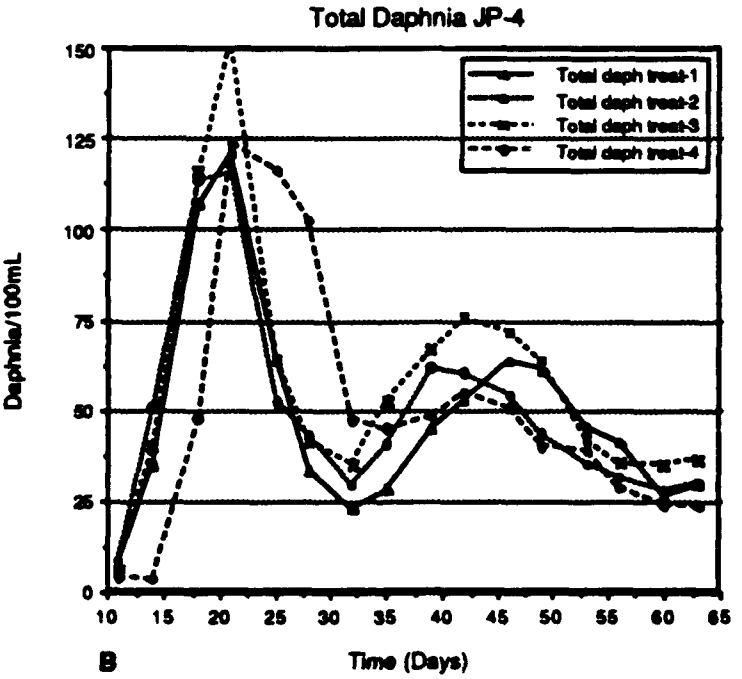


FIG. 2--Daphnid population dynamics.

The experiment using JP-4 as a toxicant (B) did not see the increase in ostracods until between days 50-55, approximately ten days later. Consequently, the total numbers of ostracods observed were not as high in the JP-4 microcosms. Note that the order of densities in the Jet-A experiment followed a dose response pattern, as did the JP-4 experiment, even with the lower total numbers. Conventional analysis did not demonstrate significance, however non-metric clustering did indicate the importance of the ostracods in determining clusters in both sets of microcosm experiments.

Philodina Population Dynamics. Philodina did not become prevalent in the microcosms until the second half of the experiment. One of the major problems was the inherent variability in the sampling and in the replicates. Organisms that reproduce rapidly can show large differences in population sizes during the course of a sampling day. Although, in the later stages of the microcosm experiments the dosed systems had a generally larger number of the rotifers, the results were not statistically significant using conventional IND plots. However, using cluster analysis, Philodina were also determined to be an important variable in defining clusters. This held true for both the Jet-A and JP-4 experiments.

Comparisons of pH dynamics of the Jet-A and JP-4 Experiments. Unlike the biotic variables, pH did reflect some of the the oscillations detected by the cluster analysis (Fig. 4). In both the Jet-A and the JP-4 experiments the highest concentrations demonstrated a statistically significant difference, determined by the interval of non-significant difference during the first 30 days of the experiment. The second oscillation, between days 45 and 50, is not as clear since only one sampling date demonstrated the statistically significant difference. Type II error becomes a concern with so many comparisons, even with the corrections incorporated into the IND plots.

Photosynthesis/Respiration Ratio. The photosynthesis/respiration ratio reflects the oscillations seen in pH and the clustering analysis for the first 30 days and then only for the Jet-A water soluble fraction. In the Jet-A experiment, a second deviation from the IND plot was noted in the period corresponding to the second oscillation, but the result is difficult to distinguish from a type II error. In the JP-4 experiment, the IND plots are large, reflecting the variance in those sampling days. As an "emergent property", it is not clear if the P/R ratio provides any more information in this experiment than the clustering based upon the biotic components.

Oscillations in Community Dynamics Observed in both the Jet-A and the JP-4 Experiments. The Jet-A and the JP-4 SAM experiments both displayed a series of oscillations; revealed by the three clustering techniques employed in the analysis (Fig. 5). The first oscillation, as defined by Cosine Distance common to each experiment, is due to the

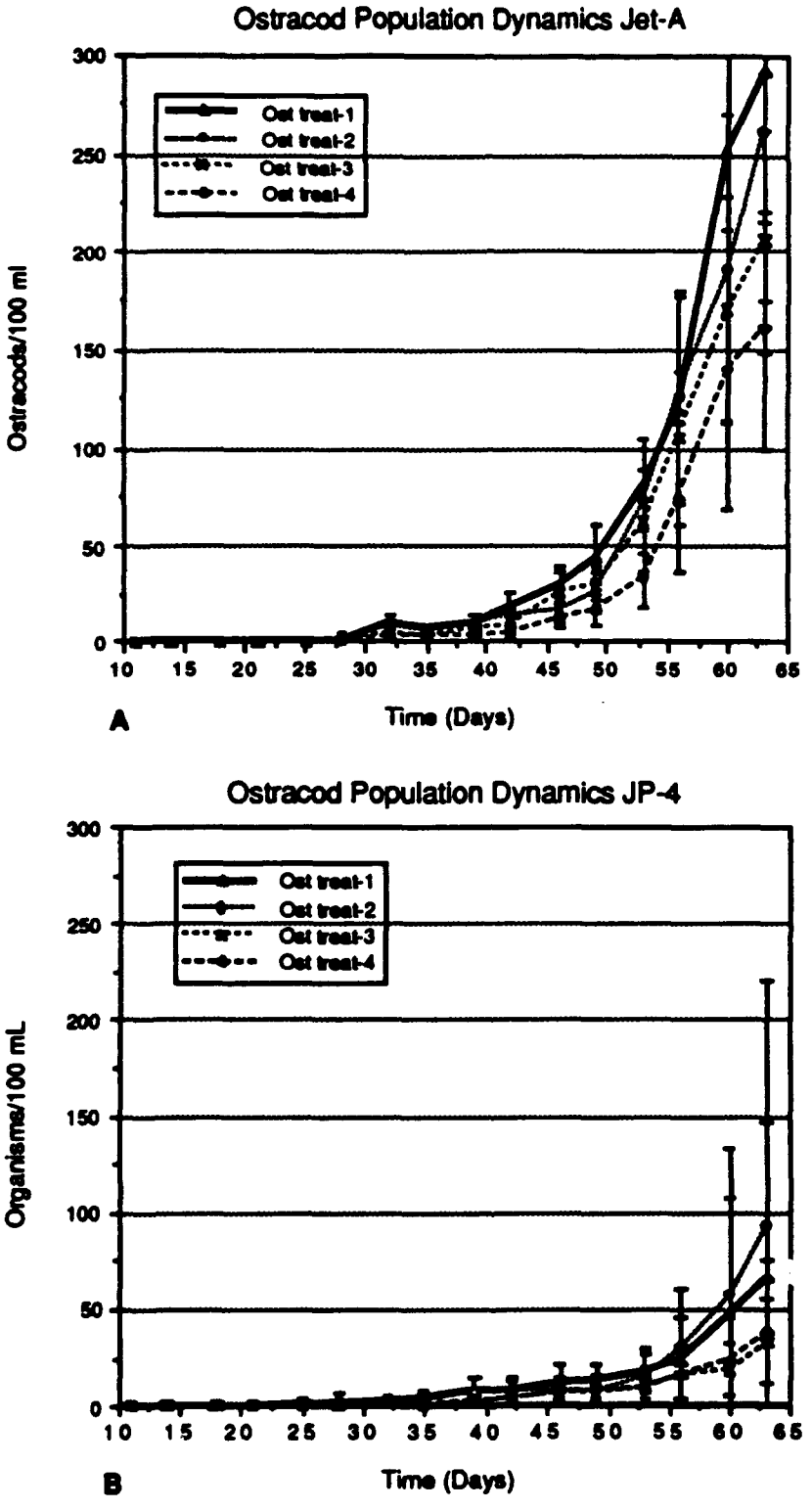
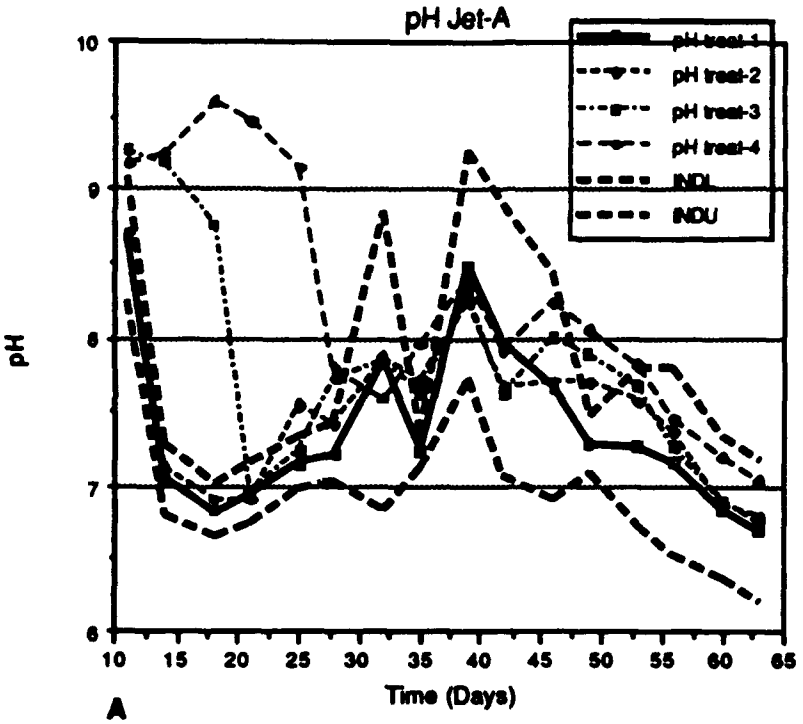


FIG. 3--Ostracod population dynamics.



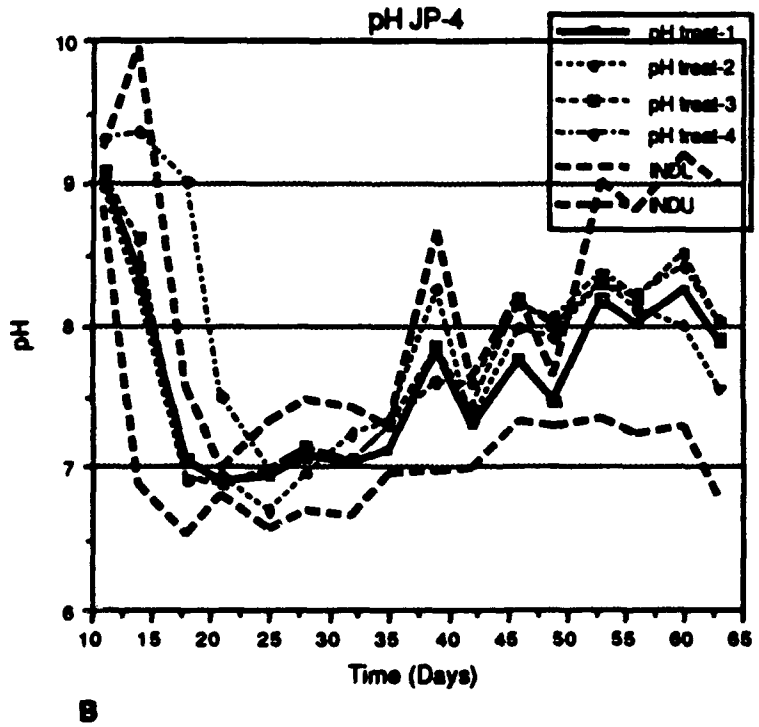
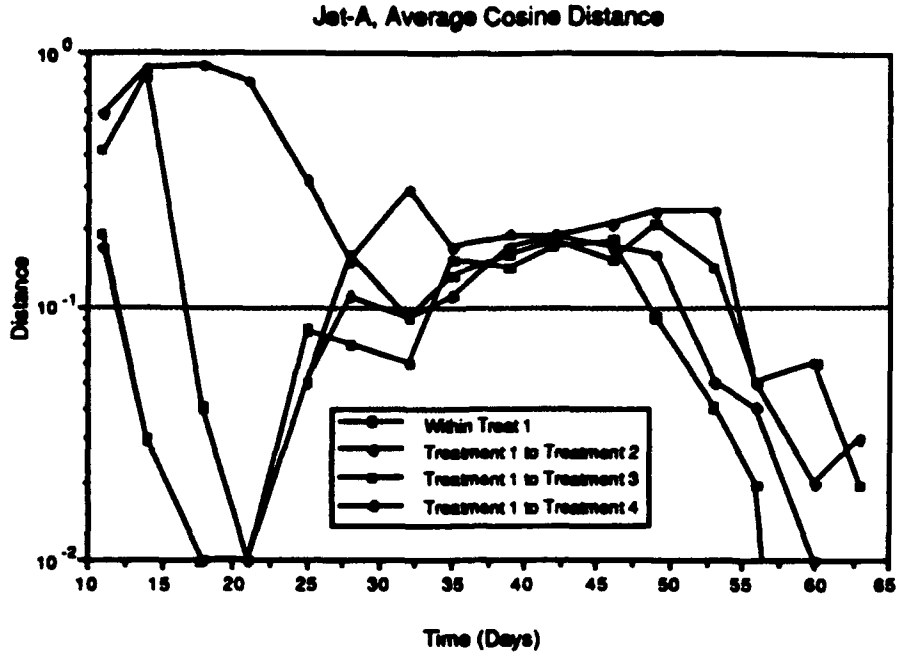


FIG. 4--Comparisons of pH during the SAM studies.



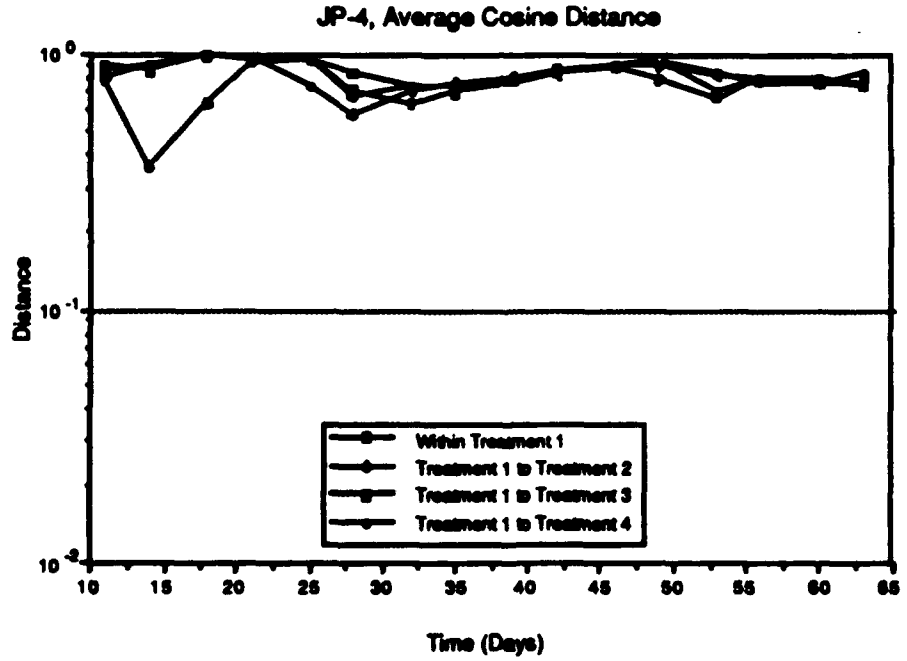


FIG. 5--Cosine Distances of the Treatment 1 to the dosed treatments in the Jet-A and the JP-4 SAMs.

interaction of the daphnid population and the algae. The result is statistically significant, as determined by the goodness-of-fit confidence level, graphed by day in Fig. 6. In both experiments, the oscillation is within the first 30 days of the SAM time-line. Interestingly, the magnitude of the first oscillation, as determined by Cosine Distance, is less in the JP-4 experiment, possibly reflecting the reduced acute and chronic toxicity of the mixture.

A second series of oscillations, as measured by Cosine Distance, occur in the last thirty days of each experiment. Again the oscillations are statistically significant.

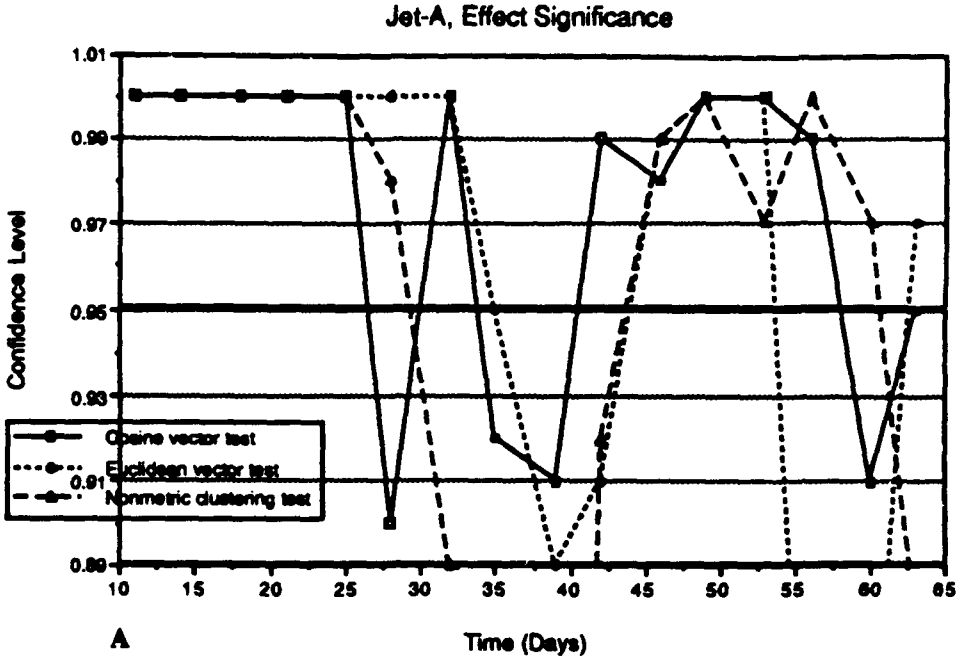
TABLE 2. Variable ranking by success in determining clusters as defined by nonmetric clustering. Variables such as Ankistrodesmus and the Daphnia classes ranked highly in the course of this study. However, reliance on any particular organism or a small combination of variables would inadequately describe the dynamics of the system.

Jet-A		JP-4	
Variable	Ranked	Variable	Ranked
Ankistrodesmus	12	Chlorella	8
M. Daphnia	11	S. Daphnia	8
Chlorella	9	Ankistrodesmus	6
Scenedesmus	7	Scenedesmus	5
S. Daphnia	6	Philodina	5
L. Daphnia	5	M. Daphnia	4
Ostracod	4	Lyngbya	4
Philodina	4	L. Daphnia	3
Selenastrum	4	Ostracod	3
Lyngbya	3	Selenastrum	3
Ulothrix	1		

The participants in the community that contribute to these oscillations are slightly different judging by the table of important variables (Table 2). Unfortunately, the length of the SAM protocol is not sufficient to conduct an analysis of the period and amplitude of the oscillations. Another complication in examining the results is the difficulty in making direct comparisons between experiments. Although the Cosine Distance may be the same, the orientation of the angle can be quite different.

DISCUSSION

First, the apparent recovery or movement of the dosed systems towards the reference or treatment 1 case may be an artifact of our measurement systems that allow the n-dimensional data to be represented in a two dimensional system. In an n-dimensional sense, the systems may be moving in opposite directions and simply pass by similar coordinates during certain time intervals. Positions may be similar but the n-dimensional vectors describing the movements of the systems can be very different. A representation of these dynamics is presented in Fig. 7. The two systems intersect, although the vectors are quite different.



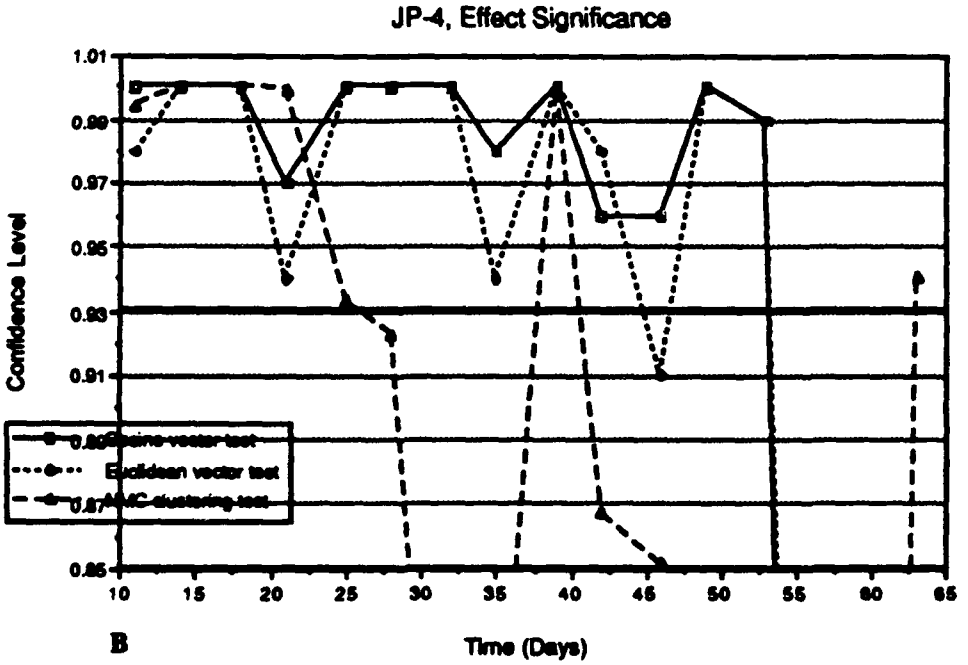


FIG. 6--Significance of the association analysis of the 4 Treatments in the Jet-A and the JP-4 SAMs.

The apparent recoveries and divergences may also be artifacts of our attempt to choose the best means of collapsing and representing n-dimensional data into a two or three dimensional representation. In order to represent such data it is necessary to project n-dimensional data into three or less dimensions. As information is lost as the shadow from a cube is projected upon a two dimensional screen, a similar loss of information can occur in our attempt to represent n-dimensional data. Not every divergence from the reference treatment may have a cause directly related to it in time. Differentiating those events from those due to degradation products or other perturbations is challenging.

Not only may system recovery be an illusion, but there are strong theoretical reasons that seem to indicate that recovery to a reference system may be impossible or at least unlikely. In fact, systems that differ only marginally in their initial conditions and at levels probably impossible to measure are likely to diverge in unpredictable manners. May and Oster (1978) in a particularly seminal paper investigated the likelihood that many of the dynamics seen in ecosystems that are generally attributed as chance or stochastic events are in fact

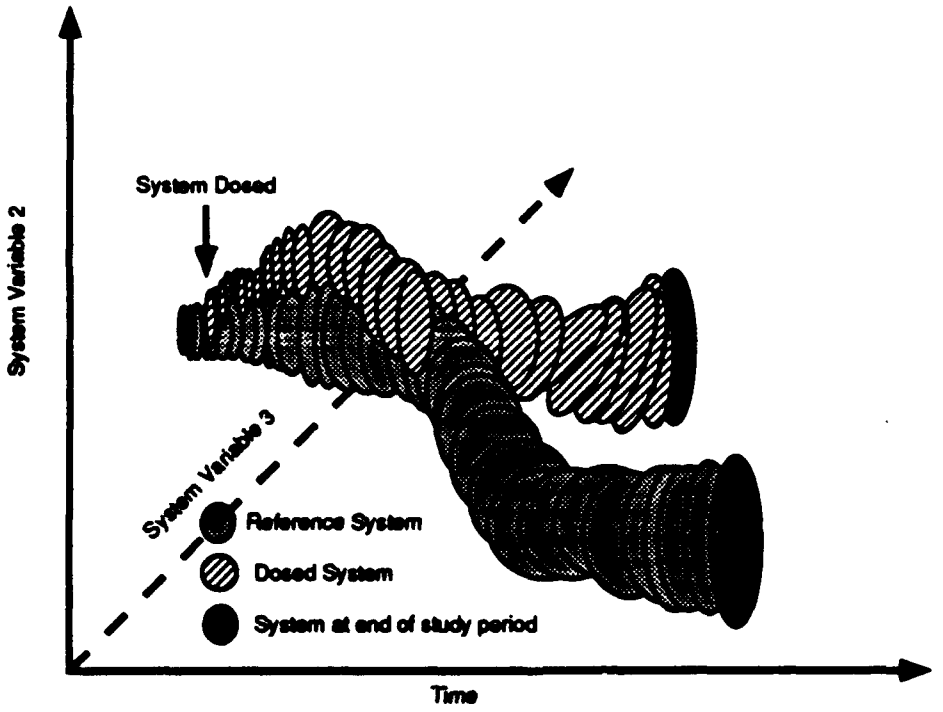


FIG. 7--Visualization of ecosystem dynamics to reflect a possible interpretation of the impacts of the jet fuels.

deterministic. In fact, simple deterministic models of populations can give rise to complex dynamics. Using equations resembling those used in population biology, bifurcations occur resulting in several distinct outcomes. Eventually, given the proper parameters, the system appears chaotic in nature although the underlying mechanisms are completely deterministic. Obviously, biological systems have limits, extinction being perhaps the most obvious and best recorded. Another ramification is that the noise in ecosystems and in sampling may not be the result of

a stochastic process but the result of underlying deterministic, but chaotic relationships.

These principals also apply to spatial distributions of populations as recently reported by Hassell et al. (1991). In a study using host-parasite interactions, a variety of spatial patterns were developed using the Nicholson-Bailey model. Host-parasite interactions demonstrated dynamics ranging from static 'crystal lattice' patterns, spiral waves, chaotic variation, or extinction with the appropriate alteration of only three parameters within the same set of equations. The deterministically determined patterns could be extremely complex and not distinguishable from stochastic environmental changes.

Given the perhaps chaotic nature of populations it may not be possible to predict species presence, population interactions, or structural and functional attributes. Kratz et al. (1987) examined the spatial and temporal variability in zooplankton data from a series of five lakes in North America. Much of the analysis was based on limnological data collected by Brige and Juday from 1925 to 1942. Copepods and cladocera, except *Bosmina*, exhibited larger variability between lakes than between years in the same lake. Some taxa showed consistent patterns among the study lakes. They concluded that the controlling factors for these taxa operated uniformly in each of the study sites. However, in regards to the depth of maximal abundance for calanoid copepods and Bosmina, the data obtained from one lake had little predictive power for application to other lakes. Part of this uncertainty was attributed to the intrinsic rate of increase of the invertebrates with the variability increasing with a corresponding increase in r_{max}. A high r_{max} should enable the populations to accurately track changes in the environment. Katz et al suggest that these taxa be used to track changes in the environment. Unfortunately, in the context of environmental toxicology, the inability to use one "reference" lake to predict the non-dosed population dynamics of these organisms in another eliminates comparisons of the two systems as measures of anthropogenic impacts.

A better strategy may be to let the data and a clustering protocol identify the important parameters in determining the dynamics of and impacts to ecological systems. This approach has been recently suggested independently by Dickson et al. (1992) and Matthews and Matthews (Matthews et al. 1991b, Matthews and Matthews 1991). This approach is in direct contrast to the more usual means of assessing anthropogenic impacts. One classical approach is to use the presence or absence of so called indicator species. This assumes that the tolerance to a variety of toxicants is known and that chaotic or stochastic influences are minimized. A second approach is to use hypothesis testing to differentiate metrics from the systems in question. This second approach assumes that the investigators know *a priori* the important parameters to measure. Given that in our relatively simple SAM systems that the important parameters in differentiating non-dosed from dosed systems change from sampling period to sampling period, this assumption can not be made. Classification approaches such as nonmetric clustering or the canonical correlation methodology developed by Dickson et al, eliminates these assumptions.

These results presented in this report and by others reviewed above and the implications of chaotic dynamics suggest that reliance upon any one variable or an index of variables may be an operational convenience that may provide a misleading representation of pollutant effects and associated risks. The use of indices such as diversity and the Index of Biological Integrity have the effect of collapsing the dimensions of the descriptive hypervolume. Indices, since they are

composited variables, are not true endpoints. The collapse of the dimensions that are composited tends to eliminate crucial information, such as the variability in the importance of variables. The mere presence or absence and the frequency of these events can be analyzed using techniques such as nonmetric clustering that preserve the nature of the dataset. A useful function was certainly served by the application of indices, but the new methods of data compilation, analysis and representation derived from the Artificial Intelligence tradition can now replace these approaches and illuminate the underlying structure and dynamic nature of ecological systems.

The implications are important. Currently, only small sections of ecosystems are monitored or a heavy reliance is placed upon so called indicator species. These data suggest that to do so is dangerous, may produce misleading interpretations resulting in costly error in management and regulatory judgments. Much larger toxicological test systems are currently analyzed using conventional statistical methods on the limit of acceptable statistical power. Interpretation of the results has proven to be difficult, if not confusing. Application of the approach and tools that proved successful in revealing the complex dynamics of these small microcosms should prove useful in analyzing larger toxicological test systems and field research.

CONCLUSIONS

(1) In both of the experiments, multiple oscillations of the dosed treatment groups away from the reference treatment were observed using multivariate statistics. The first oscillation is due to the differential impact of the WSF of the jet fuels to the algae-daphnid population dynamics. The following oscillations, although statistically significant and seen in both experiments, is not as clear cut.

The divergence of the second oscillation may be due to two separate mechanisms.

(a) A fluctuation due to the initial stress has occurred, but in such a fashion that an incompletely dampened oscillation repeats. There has been no fundamental alteration in the functioning of the ecosystem, and the oscillations are a result of the inherent time lags and stochastic factors governing the dynamics of the system.

(b) A fundamental aspect of the ecosystem has been altered so that the repeated oscillations reflect the persistence of the impact. An alteration in the detritus quality or in the community involved in the recycling of detritus may have long term impacts as other nutrients become limiting in the system. Nutrients are at low levels during the second 30 days of a typical SAM experiment. This possibility could include a fundamental and long lasting effect upon the system, contrary to the first mechanism.

(2) A combination of multivariate analyses appear to be useful and illuminating in assessing the long term dynamics of these systems. Each has strengths that make multivariate analysis a strong methodology with powerful advantages to conventional univariate methods.

(3) Although simple systems, the SAM experiments exhibits complex dynamics and behaviors. The protocol results in a persistent system with good replicability within an experiment, even with complex species interactions.

(4) Techniques that allow the reduction and visualization of even these relatively simple multispecies toxicity tests should contribute to our understanding of system dynamics and improve hazard assessment.

ACKNOWLEDGEMENTS

This research was funded by United States Air Force Office of Scientific Grant No. AFOSR-91-0291 DEF.

REFERENCES

ASTM D2887. Standard Test Method for Boiling Range Distribution of Petroleum Fractions by Gas Chromatography, 1988 Annual book of ASTM Standards, Vol. 5.02, pp 506-513. American Society for Testing and Materials, Philadelphia.

ASTM D3710. Standard Test Method for Boiling Range Distribution of Gasoline and Gasoline Fractions by Gas Chromatography, 1988 Annual Book of ASTM Standards, Vol. 5.03, pp 78-88. American Society for Testing and Materials, Philadelphia.

ASTM E 1366-91. 1991. Standard Practice for the standardized aquatic microcosm: fresh water, Vol 11.04. pp 1017-1051. American Society for Testing and Materials, Philadelphia.

Conquest, L.L. and Taub, F.B. 1989. Repeatability and reproducibility of the Standard Aquatic Microcosm: Statistical properties. In U.M. Cowgill and L.R. Williams, eds., *Aquatic Toxicology and Hazard Assessment: 12th Volume, ASTM STP 1027.* American Society for Testing and Materials, Philadelphia, PA, pp. 159-177.

Crow, M.E. and Taub, F.B. 1979. Designing a microcosm bioassay to detect ecosystem level effects. *Intern. J. Environmental Studies.* 141-147.

Dickson, K.L., Waller, W.T., Kennedy, J.H. and Ammann, L.P. 1992. Assessing the relationship between ambient toxicity and instream biological response. *Env. Tox. Chem.* 11:1307-1322.

Hassell, M.P.H, Comins, N. and May, R.M. 1991. Spatial structure and chaos in insect population dynamics. *Nature* 353:255-258.

Johnson, A.R. 1988a. Evaluating ecosystem response to toxicant stress: a state space approach. *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP* 971, W.J. Adams, G.A. Chapman and W.G. Landis Eds., American Society for Testing and Materials, Philadelphia, pp. 275-285.

Johnson, A.R. 1988b. Diagnostic variables as predictors of ecological risk. *Environmental Management* 12:515-523.

Katz, T.K., Frost, T.M. and Magnuson, J.J. 1987. Inferences from spatial and temporal variability in ecosystems: Long-term zooplankton data from lakes. *Amer. Nat.* 129:830-846.

Kersting, K. 1984. Development and use of an aquatic micro-ecosystem as a test system for toxic substances. Properties of an aquatic micro-ecosystem IV. *Int. Revue ges. Hydrobiol.* 69:567-607.

Kersting, K. 1985. Properties of an aquatic micro-ecosystem V. Ten years of observations of the prototype. *Verh. Internat. Verein. Limnol.* 22:3040-3045.

Kersting, K. 1988. Normalized ecosystem strain in micro-ecosystems using different sets of state variables. *Verh. Internat. Verein. Limnol.* 23:1641-1646.

Kindig, A.C., Loveday, L.C. and Taub, F.B. 1983. Differential sensitivity of new versus mature synthetic microcosms to streptomycin sulfate treatment. *Aquatic Toxicology and Hazard Assessment: Sixth Symposium, ASTM 802.* W.E. Bishop, R.D. Cardwell and B.B. Heidolph Eds. American Society for Testing and Materials, Philadelphia, pp. 192-203.

Landis, W.G., Haley, M.V. and Chester, N.A. 1993. The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In W.G. Landis, J. Hughes and M. Lewis., eds., *In press, Environmental Toxicology and Risk Assessment: First Volume, ASTM STP -1179.* American Society for Testing and Materials, Philadelphia, PA.

Landis, W.G., Matthews, R.A., Markiewicz, A.J., Shough, N.J. and Matthews, G.B. 1993. Multivariate analysis of the impacts of turbine fuel using a standard aquatic microcosm toxicity test. *J. Env. Sci.* In Press.

Matthews, G.B. and Hearne, J. 1991. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 13:175-184.

Matthews, G.B. and Matthews, R.A. 1991. A model for describing community change. In *Pesticides in Natural Systems: How Can Their Effects Be Monitored? Proceeding of the Conference,* Environmental Research Laboratory/ORD, Corvallis, OR, EPA 9109/9-91/011.

Matthews, G.B., Matthews, R.A. and Hachmoller, B. 1991b. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences.* 48:2184-2190.

Matthews, R.A., Matthews, G.B. and Ehinger, W.J. 1991a. Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modeling.* 53:167-187.

May, R.M. and Oster, G.F. 1978. Bifurcations and dynamical complexity in simple ecological models. *Amer. Nat.* 110:573-599.

Taub, F.B. 1969. Gnotobiotic models of freshwater communities.*Verh Internat. Verein. Limnol.* 17, 485-496.

Taub, F.B. 1976. Demonstration of pollution effects in aquatic microcosms. *Intern J. Environmental Studies.* 10, 23-33.

Taub, F.B. 1988. Standardized aquatic microcosm - development and testing. *Aquatic Ecotoxicology* II.

Taub, F.B. 1989. Standardized aquatic microcosms. *Environm. Sci. Technol.* 23:1064-1066.

Taub, F.B., Kindig, A.C. and Conquest, L.L. 1987. Interlaboratory testing of a standardized aquatic microcosm. In W.J. Adams, G.A. Chapman and W.G. Landis, eds., *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP 971.* American Society for Testing and Materials, Philadelphia, PA, pp. 385-405.

Taub, F.B., Kindig, A.C., Conquest, L.L. and Meador, J.P. 1988. Results of the interlaboratory testing of the Standardized Aquatic Microcosm protocol. In G. Suter and M. Lewis, eds., *Aquatic Toxicology and Hazard Assessment: Eleventh Symposium, ASTM.* American Society for Testing and Materials, Philadelphia, PA.

Taub, F.B. and Crow, M.E. 1978. Loss of a critical species in a model (laboratory) ecosystem. *Verh. Internat. Verein. Limnol.* 1270-1276.

Taub, F.B. , Crow, M.E. and Hartmann, H.J. 1980. Responses of aquatic microcosms to acute mortality. *Microcosms in Ecological Research.* Giesy, J.P. Jr. Technical Information Center, U. S. Department of Energy. Washington, D.C., 513-535.

Westendorf, R.G. 1986. Performance aspects of volatile organics analysis by purge and trap capillary column gas chromatography with flame ionization detectors. Tekmar Technical Papers, Tekmar Company, Cincinnati, Ohio.

Geoffrey B. Matthews[1], Robin A. Matthews[2], and Wayne G. Landis[2]

Nonmetric Clustering and Association Analysis: Implications for the Evaluation of Multispecies Toxicity Tests and Field Monitoring[3]

REFERENCE: Matthews, G. B., Matthews, R. A., and Landis, W. G., **"Nonmetric Clustering and Association Analysis: Implications for the Evaluation of Multispecies Toxicity Tests and Field Monitoring,"** Environmental Toxicology and Risk Assessment-Third Volume, ASTM 1218, Jane S. Hughes, Gregory R. Biddinger, and Eugene Mones, Eds. American Society for Testing and Materials, Philadelphia, 1994.

ABSTRACT: Many techniques developed by computer scientists in the field of artificial intelligence (AI) are currently being used as standard, state-of-the-art technology. These techniques have repeatedly proven their value and validity in medicine, geology, agronomy, and astronomy. We present here an analysis tool for multispecies data based on nonmetric clustering, an AI technique developed specifically to aid in the interpretation of complex ecological data sets. This technique uses AI search to find an appropriate and meaningful characterization of a multivariate system. After appropriately characterizing the system in this fashion, the relationship between this characterization of the system and the critical environmental variables (pollution, toxicity, *etc.*) can be quantitatively analyzed to aid in the assessment of the effects of the environment on the system. *A priori* endpoints or indices are not necessary; the data are allowed to determine the variables that best separate treatment from controls.

---

[1]Professor, Computer Science Department, Western Washington University, Bellingham, WA 98225.

[2]Professor, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA 98225.

[3]This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

We have now tested this methodology over a series of multispecies toxicity tests using a variety of stressors. During the initial blind testing the methodologies could pick treatment groups with high accuracy. When knowledge of treatment group is available, oscillations in the similarity of the treatments to the controls are apparent.

Much recent debate in toxicological studies has focussed on appropriate endpoints for multispecies toxicity tests and biomonitoring schemes. We suggest that the search for endpoints appropriate to the entire field of toxicity testing is a fruitless search. We recommend instead an approach that standardizes the common sense approach: different situations, even within a single experiment, call for different endpoints. Typically, the toxicologist, if called upon for an expert opinion, will examine multivariate data, and extract from that data a few critical species. The behavior of these species will give an adequate (though perhaps not complete) picture of the toxic effects. Which species are selected, and whether it is their mortality, behavior, or biomass that is important, will always vary from case to case. We call, therefore, for more research into the automation of the process typically performed by the expert. The selection of species, as well as other parameters, as significant for a particular experiment or field study, can be done automatically by computer algorithms. To be blind to the utility of these tools in the field of toxicology is to work by hand, over and over again, problems which could be solved in a twinkling with their aid.

KEYWORDS: artificial intelligence, ecotoxicology, statistics, expert systems, multispecies tests, field monitoring

Introduction

Modern ecotoxicology often assumes that ecosystem level functional indices are desirable. Such an index would tell us what numbers to measure, which mathematical formulae to use on them, and the position of the cutoff point between healthy systems and troubled ones. This would introduce ''objectivity'' into what is now done with intuitive assessments. However, the state of an ecological community is an inherently complex object, and probably cannot be captured on by a few linear indices. Remaining with the traditional human ''best judgement,'' on the other hand, also has problems, such as subjectivity, prejudice, and the difficulty of comprehending the innate complexity of ecological systems. Fortunately, there is a middle ground for dealing with complex systems, between mathematical formulae and intuition. The middle ground is provided by computerized data exploration, using tools from artificial intelligence, pattern recognition, and scientific visualization. In other scientific domains, such as medicine, astrophysics, particle physics, meteorology, and geology, such tools have been in widespead use for years. The key to their success is that the human expert and the software tool are partners in the exploration of the data. The computer by itself, of course, has no semantic understanding of the data. But, equally, the unaided human may be blind to the many patterns implicit in the data.

Much of the work in computer-aided data exploration, however, has the wrong focus for ecotoxicology. Data sets generated, for example, by meteorological models of a thunderstorm, typically have millions of data points densely scattered through a well-defined three-dimensional model. The complexity is in the sheer number of data points and their interactions. In ecologically interesting situations, on the other hand, only a few dozen or hundred data points are in hand, from widely separated places in space and time, and each point records data on dozens or hundreds of species. This results in a relatively small number of points scattered through the huge volume of n-dimensional space (where n is the number of different species counted). Even a modest number of dimensions raises severe problems for conventional analysis techniques, and

human intuition. For example, if some large number of points is scattered uniformly over a 10-dimensional hypersphere with radius one, then a hypersphere inside, of radius 3/4, will contain only 5% of the points. Clearly, sampling 10 or higher dimensional space can miss important things. Further, a lot of the time data points are missing, or incomplete.

The nature of the problem is that usually we have too many dimensions. Ten or twenty sampling points with, perhaps, fifty species, is underdetermined. There is no way to draw meaningful conclusions about the nature of the community as a whole (all fifty dimensions), from the smattering of points. What is required is *data reduction*, the dimensionality of the data has to be brought down to the point where ten or twenty points can tell us something. One methodology for this is based on *projections* of the data, such as factor analysis, principal components analysis, correspondence analysis, or, more generally, projection pursuit [1]. There are many algorithms for finding good projections, and even a suggestion that *all* projections be examined in a ''grand tour'' of the data [2]. However, rotating at about 10° per second, a reasonable speed for careful observation, a grand tour of only four dimensions would take about three hours [1], and so computer-aided projections are the only real alternative.

While such projections are valuable in reducing the dimensionality of the data, they all suffer from a problem of comprehensibility. Since arbitrary linear and nonlinear transformations of the data matrix are allowed, the meaning of the resulting two-dimensional projection can be obscure, and difficult for grasp intuitively.

Another possibility for data reduction is provided by careful experimental design. Measuring too many parameters can be just as misleading as measuring too few. The problem with this approach, however, is that in many circumstances the correct parameters are not known in advance. Indeed, our understanding of ecological systems and their response to toxic stress is still in its infancy. What is needed is a tool to help the analyst in finding important and significant parameters in new situations.

The tradition of machine learning (ML) within artificial intelligence, has been addressing these

problems for some time. The goal of an ML system is, not only to identify patterns in the data, but to come up with an efficient and intuitive characterization of them. Efficient and intuitive, in this context, imply that the characterization is not overly complex, that it uses simple logical combinations of descriptions rather than mathematical formulae, and that it is expressed in terms of attributes that are not contrived. This has been formulated as the *comprehensibility postulate*:

> The results of computer induction should be symbolic descriptions of given entities, semantically and structurally similar to those a human expert might produce observing the same entities. Components of these descriptions should be comprehensible as single ''chunks'' of information, directly interpretable in natural language, and should relate quantitative and qualitative concepts in an integrated fashion [3].

It is the primary failing of traditional statistical approaches, as well as the ''neural net'' approach, to solving ML problems that they ignore the comprehensibility postulate.

In this paper, we present nonmetric clustering, a specialization of ML, faithful to the comprehensibility postulate, which we have been employing successfully on a wide variety of ecosystems. After its details are explained, some consequences for environmental policy making are outlined.

Machine Learning

As a simple example, consider the data in Table 1 [4]. In this set, we are given three ''positive'' individuals and five ''negative'' individuals and their characteristics on three attributes. The problem is to come up with a means of distinguishing the ''positives'' from the ''negatives'' based on height, hair color, and eye color. There are many possible ways of distinguishing them, but one nice one might be:

> Positives either have red hair, or blond hair and blue eyes.
> Negatives either have dark hair, or blond hair and brown eyes.

Height	Hair	Eyes	Class
short	blond	blue	+
tall	blond	brown	−
tall	red	blue	+
short	dark	blue	−
tall	dark	blue	−
tall	blond	blue	+
tall	dark	brown	−
short	blond	brown	−

Table 1: Data set problem for identification and characterization.

There are several things to notice about this characterization of the positives and negatives.

First, the data are both categorical and numeric. The beauty of ML approaches to these problems is that they apply equally well to either kind of data. To make a regression, or linear discriminant, categorical data would have to be numerically coded somehow. A loglinear model can be used on categorical data, but then the numeric data would have to be fit in. In an ML approach, numeric attributes, such as height, are simply recoded into a number of discrete bins, such as small, medium, and large. Such categories can be as fine or as coarse as desired, and in all events are more comprehensible than an uninterpreted number.

Second, not all the original attributes are used in the description. Height, it turns out, is superfluous, and is omitted from the description.

Third, compound descriptions are created using logical operations, ''and'' ''or'' and ''not'', rather than mathematical formulae. A linear discriminant, for example, describes by adding up numbers and then determining if the result is greater or smaller than some cutoff point. The logical descriptions are much more natural and intuitive for humans, and lead to understanding of the data in a way that mathematical combinations cannot.

Fourth, even with only three attributes and eight points, there are a lot of different logical descriptions that have to be considered to get the best one (or even a good one). With real data sets the combinatorial complexity of finding a description would rapidly swamp a

C1	A	B	C	D	E	F	C2
+	1	2	1	2	2	1	+
+	1	2	2	1	1	2	−
+	1	2	1	2	2	1	+
−	2	1	2	1	1	2	−
−	2	1	1	2	2	1	+
−	2	1	2	1	1	2	−

Table 2: Synthetic data for nonmetric clustering, with two possible clusterings.

human investigator. A computer aid is essential.

Fifth, no artificial attributes are used. The use of ''indices'' or ''ordination'' techniques attempts to introduce a new attribute, defined mathematically in terms of the original ones, and then use the values of these indices or components to describe the classes. The ML description uses the same attributes (height, hair, and eyes) that were used in the design of the sampling program, and thus, the description of the classes will have direct meaning to the investigator, without the need to learn a new vocabulary. Such descriptions, which use simple logical combinations of the original attributes, are called ''conceptual'' descriptions [5].

Nonmetric Clustering

Nonmetric clustering (NMC) is an ML tool designed to search for conceptual descriptions of ecological data sets. The NMC methodology has been implemented in a computer program called Riffle [6]. Unlike the simple example above, Riffle does not work from a preexisting set of class labels (such as + and −). Given a data set, Riffle attempts to two things simultaneously: Group the points into clusters (classes), and find the simplest possible conceptual description of those clusters. Since the points are not previously assigned to classes, Riffle is free to give the points any class label at all. However, the class labels must be such that they can be simply captured in a conceptual description, based on the original attributes (measured parameters), and, further, such that they, in turn, capture as much information as possible about the original attributes.

Consider the synthetic data in Table 2, where six points have been sampled for six attributes. One

potential clustering, denoted C1, has two simple conceptual descriptions, each based on a single attribute, either A or B. C, D, E and F can be regarded as superfluous for this clustering. Another potential clustering, denoted C2, also has simple characterizations, but in terms of attributes C, D, E, and F, with A and B as superfluous. While both clusterings have simple conceptual descriptions, C2 should be preferred because it captures more information about the points than C1. One way to express this algorithmically is that there are *more* good conceptual descriptions of the classes in C2 than there are of the classes in C1. The computer program Riffle will prefer C2 to C1 for this reason.

To find the best clustering possible, for a given data set, the algorithm works by examining a great number of possible clusterings, like C1 and C2, above, and numerically ranks their conceptual adequacy. All data points are repeatedly reassigned to clusters, and then the conceptual association between clusters and attributes is reevaluated. When an assignment of points to clusters is found that outranks all others, it is reported as the most natural clustering.

We will now briefly discuss how conceptual adequacy is ranked, and also make some remarks on the particular strategy used in Riffle to convert numeric to categorical variables.

Numerically ranking conceptual descriptions

To begin with, assume all attributes are categorical. Nonmetric clustering measures the association between a clustering (which, itself, is a categorical variable) and another categorical variable by means of a contingency table test. A frequency table of cluster-number *vs.* categorical-value is set up, and the number of data points in each cell is counted in order to measure the association between cluster and variable. The most famous contingency table test is probably the χ^2 test, but the χ^2 test has some undesirable properties when it comes to interpretation and comprehensibility. Nonmetric clustering uses Guttman's λ to measure the association in the table [7, 8, 9, 10].

Guttman's λ is a measure defined on the basis of ''optimal predictions''. Consider, for instance, the

	A1	A2	A3
B1	5	3	1
B2	1	4	2
B3	7	0	5

Table 3: A contingency table to illustrate calculation of Guttman's λ.

contingency table represented in Table 3. Twenty-eight individuals have been sampled, and their values on attributes A and B have been tabulated. For concreteness, A can be regarded as ''height'' and B as cluster-number. A larger sample size would always be desirable, but we have no recourse other than to regard the proportion of points found in any cell as the best estimate of the probability of finding a new point also to be in that cell. Now suppose we need to predict which value on attribute B a new sample is likely to have. In the absence of any further information, there are nine B1's, seven B2's, and twelve B3's, so we would guess B3, and expect to be right about 12 out of 28 times, giving us an error expectation of 16 out of 28, or about 57%. We will call this the *absolute error rate of B*. Now, however, suppose we are given a new data point, and are told its value for attribute A. How will we predict B, and what will our expected error rate be when conditioned on this knowledge? Well, 13/28 of the time the new point will be A1, and we should then guess B3, and expect to be right 7/13 of the time. Similarly, 7/28 of the time it will be A2, and we will guess B2, and be right 4/7 of the time, and 8/28 of the time it will be A3, we guess B3, and are right 5/8 of the time. Predictions of B conditioned on A, then, should be correct (13/28)(7/13) + (7/28)(4/7) + (8/28)(5/8) $\approx$ 57% of the time, and the *error rate of B conditioned on A* is 43%. The *reduction in error* is $57-43$, and the *proportional reduction in error* is $(57-43)/53 \approx 26\%$. In comprehensible terms, we expect to be wrong about 26% fewer times if we know A. The proportional reduction in error when predicting A conditioned on B can be computed similarly. The absolute error rate of A is $(28-13)/28 \approx 54\%$, the error rate of A conditioned on B is 1 - [(9/28)(5/9) + (7/28)(4/7) + (12/28)(7/12)] $\approx$ 43%, and the proportional reduction in

error is $(54-43)/54 \approx 20\%$. Each of these proportional reductions in error is a measure of how well the knowledge of one attribute aids in the prediction of the other. A symmetric measure of association can be obtained by simply averaging the two conditioned measures, giving the symmetric λ, of 23%.

Formally, λ can be defined as follows. For a contingency table with proportional entries ρ_{ab}, let $\rho_{\cdot b} = \sum_a \rho_{ab}$, $\rho_{a\cdot} = \sum_b \rho_{ab}$, $\rho_{am} = \max_b \rho_{ab}$, $\rho_{mb} = \max_a \rho_{ab}$, $\rho_{\cdot m} = \max_b \rho_{\cdot b}$, and $\rho_{m\cdot} = \max_a \rho_{a\cdot}$. Then the reduction in error of b with respect to a is

$$\lambda_b = \frac{\sum_a \rho_{am} - \rho_{\cdot m}}{1 - \rho_{\cdot m}}$$

and the reduction in a with respect to b is, similarly,

$$\lambda_a = \frac{\sum_b \rho_{mb} - \rho_{m\cdot}}{1 - \rho_{m\cdot}}$$

The symmetric form, which averages these two cases, becomes

$$\lambda = \frac{1}{2} \frac{\sum_a \rho_{am} + \sum_b \rho_{mb} - \rho_{\cdot m} - \rho_{m\cdot}}{1 - \frac{1}{2}(\rho_{\cdot m} + \rho_{m\cdot})}$$

Obviously, the more strongly two attributes are associated, the higher the value of λ, and *vice versa*. Some other properties of λ [7] are:

- λ lies between 0 and 1, inclusive, except when the entire population lies in a single cell of the table, in which case it is indeterminate.
- λ is 1 if and only if all the population is in cells no two of which are in the same row or column.
- Independence is sufficient, but not necessary, for λ to equal 0.
- λ is unchanged by permutations of rows or columns.

We have found λ to be an excellent measure of qualitative association, in that it accords well with intuition and is much more ''stable'' than χ^2 [11]. Using λ to calculate the association between cluster-numbers and categorical attribute values is faithful to the comprehensibility postulate: an attribute is a good

description of a clustering if knowledge of the attribute helps predict cluster, and *vice versa*.

Integrating qualitative and quantitative data

The frequency table approach works well for categorical variables, but what about numeric variables? Nonmetric clustering takes a pragmatic approach to these: if we assume that the data are going to be adequately described by a clustering into a finite number of clusters, then there are really only a finite number of values of a numeric parameter to consider, one for each cluster. All other variations in a numeric parameter can be assumed to be due to variance within the clusters. Accordingly, we can divide up the range of a numeric parameter into discrete parts. We can do this by simply choosing quantile points, but a more flexible arrangement allows the ''splits'' between categorically different values to be selected by the algorithm as it runs. How this is accomplished is illustrated in Figure 1. Here we have marked two clusters with open and filled circles, and the categorical division of two dimensions into ''high'' and ''low'' values are shown by the dividing gray lines. The point marked with an ''X'' is troublesome, as it does not fit well with either of the two clusters, and keeps us from obtaining a λ value of 1.0 for this data set. We could move the vertical line to the right, changing our division between ''long'' and ''short'', to try to include X in one cluster, but that would raise more problems by the inclusion of some points from the other cluster. Similar problems occur if we try to raise the horizontal line, changing our definition of ''light'' and ''heavy''.

The computer program Riffle will keep adjusting these split lines up and down to achieve better associations between cluster and numeric attribute. In other words, what counts as ''small'' or ''large'' can be redefined by the algorithm as it investigates the data. At the same time, the algorithm is free to reassign the points themselves to different clusters. Both of these reinterpretations of the data are tried over and over, to maximize λ. The algorithm stops when it cannot improve the association between clusters and attribute any more.

This clustering methodology has a number of advantages over traditional clustering methods:

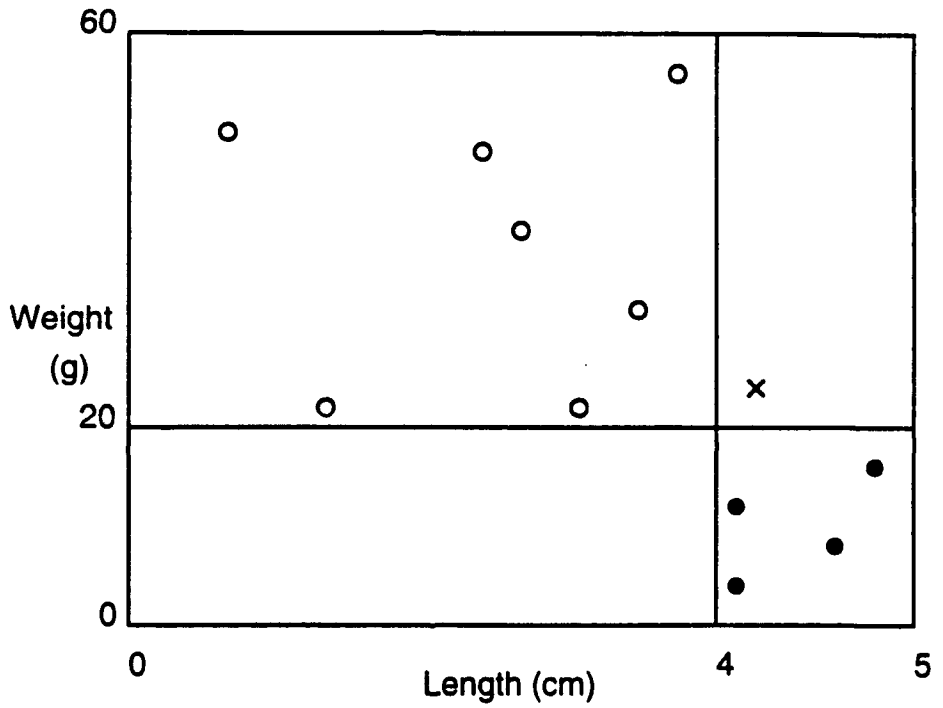


Figure 1: Twelve synthetic data points in two dimensions. Clusters are indicated by open and filled circles. Split values shown by gray lines. The point marked with an "X" cannot be included in either cluster by moving the split values without introducing further problems.

- It does not combine counts from dissimilar taxa by means of sums of squares, or other *ad hoc* mathematical techniques.
- It does not require transformations of the data, such as normalizing the variance.
- It works without modification on incomplete data sets. Since each attribute has its λ-association with the clustering evaluated independently, the fact that some points have some values for some attributes, and other points for other attributes, is irrelevant. Attributes are *not* directly combined.
- It can work without further assumptions on different data types (*e.g.*, numeric, categorical, species counts, presence/absence data, *etc.*).
- Significance of an attribute to the analysis is not dependent on the absolute size of its count. For instance, a taxon having a small total variance, such as rare taxa, can compete in importance with common taxa, and taxa with a large, random variance will not automatically be selected, to the exclusion of others.
- It provides an integral measure of ''how good'' the clustering is, *i.e.* whether the data set differs from a random collection of points, by means of the size of the λ values for each attribute.
- It can, in some cases, identify a subset of the attributes that serve as reliable indicators of the physical environment. In our research the indicator species selected by Riffle often proved to be more reliable than indicators based on a linear discriminant [12, 13].

The major disadvantage of the Riffle program is that, in order to find a clustering of the data points with the desirable qualities listed above, a massive search through thousands of potential clustering candidates is made before settling on the ''right'' one. Even after this search, there is no guarantee that Riffle finds the optimal clustering, in the sense outlined above. However, in our research, involving datasets with one or

two hundred dimensions and points, Riffle does find an excellent clustering in a reasonable amount of time. For larger datasets, supercomputers and/or more heuristic searches may be required.

Riffle has been applied successfully in a number of ecological and toxicological situations. For example, in a study of urban runoff in a small stream [12], Riffle was able to identify a community of macroinvertebrates that was associated with clean water regardless of seasonal variation. Many of the species in the community were quite rare, and would have been overlooked without the use of Riffle. In a study of jet fuel toxicity in microcosms [14], Riffle was able to identify communities of species associated with toxic dose. Over the sixty days of the experiment the communities changed, from communities distinguished predominantly by their predator/prey ratios (*Daphnia* and algae) to communities distinguished predominantly by the makeup of their detritivores (Ostracods, *etc.*). These patterns in the data suggested new hypotheses and further experiments that would not have been conceived without Riffle's aid.

Association Analysis: a Significance Test from the Clustering

If the data analyzed have natural groups, such as treatment groups or sites, a significance test can be derived from the known groups and the generated clusters. Under the null hypothesis, clusters generated from the data will have no association with the known treatment groups. Thus, if the generated clusters closely match the treatment groups, with less than one or five percent probability under the null hypothesis, then a significant effect has been found. We have used nonmetric clustering and association analysis on a variety of multivariate experiments and find it to be comparable in sensitivity to many metric tests that make more assumptions about the underlying distributions of the data [14].

Implications for Ecological and Ecotoxicological Tests

The fact that nonmetric clustering and association analysis (NCAA) adheres to the comprehensibility postulate has numerous consequences for the analysis of ecological data, and for policy. When establishing policy for mitigation or restraint, the ecologist is forced into the position of deciding what is ''good'' and

what is ''bad,'' or natural *vs.* unnatural, or pristine *vs.* polluted, or healthy *vs.* unhealthy. The development of various ecological indicators (diversity indices, indicator species, biomarkers, *etc.*) has proceeded by fits and starts, primarily because ecosystems are complex and rarely reproducible, and so a simple division into good and bad ecosystems is not feasible. Instead, each new system must be approached on its own terms, and ecological and toxicological experts must begin to understand it afresh and derive new concepts each time.

A computational induction from the data alone using ML techniques has a number of advantages.

1. Machine learning is free from prejudice. Too often natural ecologists are forced to rely on traditional indicator species, or traditional measures of diversity, rather than taking a fresh look at each new system. Machine learning software does not remember the past, although the possibility is always open to incorporate prior information by, for example, weighting the dimensions.

2. Machine learning is adaptable. There is no need to establish policy based on a few preselected species, or on one mathematical technique. A variety of techniques, and all possible species, can be incorporated into a single ML tool which will sort through them and return with an objective picture of the ecosystem based on the most influential species and the most informative tools.

3. Machine learning is interactive. Because the concepts derived by computational induction are faithful to the comprehensibility postulate, they can be examined by human experts. The machine is not a ''black box'' which must either be trusted implicitly or thrown out completely. Refinements in the ML algorithm can be visualized, based on experiments, and reincorporated into future generations of the ML computational tools.

4. Machine learning is not constrained like expert systems. Unlike expert systems, which attempt to encapsulate a particular human's expertise in a computer system, ML tools attempt to derive new

expertise, new categories and concepts, derived from the data themselves. The only constraint on an ML system is the comprehensibility postulate, requiring that all new ideas be expressible in human terms. Beyond that, anything goes.

5. Machine learning is inexpensive. One of the primary motivations behind the surge of interest in expert systems was that a computer program represents a large initial investment, but a very small marginal cost subsequently, compared to professional consultation with a human expert. ML systems, once developed, are marketed like any other software, and can be duplicated and reused, in identical form, on any site.

Because of these advantages, we can recommend a new direction in ecotoxicological policy. There is a middle ground between reliance on completely objective, simple, numerical cutoffs, on the one hand, and largely subjective, naked faith in consensus human judgement, on the other. Rather, policy must be made only after extensive interaction between human experts and their ML assistants. Without ML and the associated computational induction, the human expert cannot be sure that some important concepts not are being overlooked. The human's compromises and policies should only be made after the minimal step of consulting with an ML system. Such man-machine consultations must become part of policy, or else we are condemned to base judgements on only partial information, on oblique, narrow, and slanted views of the data. We therefore call for ecotoxicologists to review the large ML literature, and begin to establish standards for human-computer interactive analysis of ecological systems.

Future Work: Dynamic Ecosystem Change

While our system of nonmetric clustering and association analysis does well with a variety of environmental data, we are currently seeking a much-needed extension of our ideas. At present, each data set is treated statically, as an independent point in time. In reality, environmental systems are extremely sensitive to their history. What is needed is a conceptual description of ecological systems that pays

particular attention to the dynamic nature of systems over time. On the one hand, time could simply be viewed as another measured attribute; however, it is obvious that this attribute holds a special place. Time series analysis, as it is currently practiced, is frequently a univariate technique, primarily concerned with trends and cycles. What is required is a multivariate technique that makes sense of multivariate trends in patterns. One straightforward approach is to consider the state of a multivariate system as a multivariate vector, and the change over time as simply another vector connecting the state at one time with the state at another. In this view, we could define velocity, curvature, torsion, and a host of other vectors which would, in some sense, characterize the changes of the system over time. However, we must look instead for a description of change that does not violate the comprehensibility postulate. For a conceptual clustering, we must look for a *conceptual shift*, and have a concise notion of what this means. When we have decided the terms under which conceptual shifts are described, we can then build an ML tool that will assist us in our search for understanding. We believe that a conceptual shift in the character of a community or ecological system will be far more significant than any simple change in the numbers of species.

Conclusion

Machine learning promises to revolutionize the practice of environmental policy, by making the marriage of human and computer expertise a reality. We anticipate computerized ''policy assistants'' that will create an atmosphere of understanding and familiarity with the most difficult data. We have presented here, as an illustration, our own technique of nonmetric clustering and association analaysis, which we have used repeatedly in gaining deeper insights into ecological and toxicological data. The computer tools of machine learning present a new alternative to past practices, one which is at the same time more friendly and more objective, and one which will, sooner or later, be indispensible to our field.

References

[1] P. J. Huber. Projection pursuit. *Annals of Statistics*, 13:435--475, 1985.

[2] Dan Asimov. The grand tour. *SIAM Journal of Scientific and Statistical Computing*, 6:128--143, 1985.

[3] R. S. Michalski. A theory and methodology of inductive learning. *Machine Learning, An Artificial Intelligence Approach*, pages 83--134, 1983.

[4] J. Ross Quinlan. Learning efficient classification procedures and their application to chess end games. In Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell, editors, *Machine Learning, An Artificial Intelligence Approach*, pages 463--482. Morgan Kaufmann, Los Altos, California, 1983.

[5] R. S. Michalski and R. E. Stepp. Learning from observation: Conceptual clustering. *Machine Learning, An Artificial Intelligence Approach*, pages 331--363, 1983.

[6] Geoffrey Matthews and James Hearne. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175--184, 1991.

[7] L. A. Goodman and W. H. Kruskal. Measures of association for cross classifications. *Journal of the American Statistical Association*, 49:732--764, 1954.

[8] L. A. Goodman and W. H. Kruskal. Measures of association for cross classifications ii: further discussion and references. *Journal of the American Statistical Association*, 54:123--163, 1959.

[9] L. A. Goodman and W. H. Kruskal. Measures of association for cross classifications iii: approximate sampling theory. *Journal of the American Statistical Association*, 58:310--364, 1963.

[10] L. A. Goodman and W. H. Kruskal. Measures of association for cross classifications iv: simplification of asymptotic variances. *Journal of the American Statistical Association*, 67:415--421, 1972.

[11] Chanane Chen. The measurement of clustering tendency in machine learning. Master's thesis, Western Washington University, Bellingham Washington, 1992.

[12] Geoffrey B. Matthews, Robin A. Matthews, and Barnard Hachmöller. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(11):2184--2190, 1991.

[13] Robin A. Matthews, Geoffrey B. Matthews, and William J. Ehinger. Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modelling*, 53:167--187, 1991.

[14] W. G. Landis, R. A. Matthews, A. J. Markiewicz, and G. B. Matthews. Multivariate analysis of the impacts of the turbine fuel jp-4 in a microcosm test with implications for the evaluation of ecosystem dynamics and risk assessment. *Ecotoxicology*, 2, 1993.

Nonmetric Clustering: New Approaches for Ecological Data

Geoffrey Matthews Robin Matthews Wayne Landis

Western Washington University, Bellingham, WA 98225

Abstract

Ecological studies and multispecies ecotoxicological tests are based on the examination of a variety of physical, chemical and biological data with the intent of finding patterns in their changing relationships over time. The data sets resulting from such studies are often noisy, incomplete, and difficult to envision. We have developed machine learning and visualization software to aid in the analysis, modelling, and understanding of such systems. The software is based on nonmetric conceptual clustering, which attempts to analyze the data into clusters that are strongly associated with several measured parameters. Our analysis and visualization tools not only confirmed suspected ecological patterns, but revealed aspects of the data that were unnoticed by ecologists using conventional statistical techniques.

1 Introduction

Nonmetric clustering [1] is a variant of conceptual clustering in that the clustering is designed, not only to fit the data, but also to create a *simple* and *conceptual* description of the data [2]. The goal of nonmetric clustering is a partition of the data into disjoint and exhaustive subsets (the clusters) such that most of the points can be described by simple conjunctive descriptions involving some subset of the original parameters. This differs from varieties of factor analysis [3] in that a subset of the original parameters is used rather than a rotation or projection. Our implementation of nonmetric clustering (the computer program "Riffle") performs a search through the space of all partitions of the data, and all divisions of the parameters into qualitative categories (*e.g.*, "small", "medium", and "large"), and all subsets of parameters. The search terminates when it finds a clustering (partition), parameter subset, and categorical division, such that the fit to the data cannot be improved. The space of partitions and divisions is too large to be searched exhaustively and a hill-climbing algorithm is employed, using several random starting positions.

Nonmetric clustering has some advantages over conventional clustering methodologies. First, it works well with incomplete data, where several points may have missing values for a few dimensions. Second, it works equally well with categorical, ordinal, and numeric dimensions. Third, it does not require *ad hoc* modifications of the numeric dimensions, such as normalizing the variance. Fourth, it does not rely on a metric, such as the Euclidean metric, which will combine parameters by sums of squares or other mathematical methods. Fifth, it provides an integral measure of the quality of the clustering, allowing an objective choice, *e.g.*, for the right number of clusters. Finally, it has the ability to ignore noisy parameters, *i.e.* parameters with a large variance but random with respect to the overall pattern. Size of the variance is not taken into account since all values on all dimensions are merely regarded as small, medium, or large.

The clustering itself is informative, but Riffle actually provides the user with more than a traditional clustering algorithm. It also reports a list of the parameters that have a strong association with the clusters. This list, which is a subset of all of the parameters, records only those that are important or significant in relation to the patterns in the data. Parameters that vary randomly are automatically be excluded from the list. This feature has proved invaluable to ecologists. We will describe one case here [4, 5].

2 Microcosm ecotoxicology

Riffle has been successful in analyzing data from synthetic microcosms such as the Standardized Aquatic Microcosm, or SAM [6]. In the SAM, twenty-four jars of water are prepared identically with several species of algae, *Daphnia*, and other biota. The jars are divided into four treatment groups, normally

a control and three increasingly toxic doses. The jars are monitored closely for two months and population counts for all species, as well as physical/chemical parameters, are recorded every few days.

Over most days of the experiment, nonmetric clustering by Riffle can pick out the four treatment groups from the biological data alone, even when individual parameters show no significant difference among the four groups. Further, the parameters (species) that Riffle selects as associated with the clustering reveal community-level responses to toxic stress. Quite often some species will respond early in the test, and different ones later. For instance, in at least two of our experiments, the treated groups diverged significantly from the control group, and then, by about the end of the first month, "recovered" to a state indistinguishable from the control group. However, during the second month, the treatment groups again diverged from the control.

This divergence, convergence and redivergence, or *oscillation* in distance between treatment and control, was also visualized with 3d spacetime graphics, and its statistical significance was separately confirmed by a permutation test on relative multivariate metric distances within and between groups [7]. However, while visualization and confirmatory statistics were helpful in establishing the existence of the oscillation, Riffle went further and identified a different microbial community during the early and late separations, suggesting further hypotheses to ecologists about the hidden mechanisms which determine the long-term toxic impact on the community.

The oscillation indicates that, during the putative recovery period, the systems were nonetheless quite differently affected by the toxic stress, and revealed this by a later divergence. Without further insight, this might be interpreted as a chaotic effect, in which minute differences at one time can have large, nonlinear effects later. But with the added information Riffle provided, testable hypotheses about the actual mechanisms involved were derived; we are currently investigating them. Thus we can, in this case, look beyond the apparently chaotic surface to the hidden variables underneath which may be the true causes.

3 Conclusion

Our program attempts to understand multivariate data on its own terms. To this end, we have built and applied nonmetric clustering and visualization tools that reduce the dimensionality and complexity of data from multispecies communities to a manageable size. The reduced data is more interpretable by scientists, and has aided in the discovery of new hypotheses regarding community level response to toxic stress.

We believe that our success with this methodology suggests a new paradigm for toxicity testing. Traditionally, assessment of toxicity is done using one or two species deemed suitable. With machine learning tools, however, the species most suitable to assessing the effects of the toxic agent need not be specified in advance, but can be discovered automatically from within the multispecies test. This will provide at once a more sensitive, and a more realistic, test.

Acknowledgements

This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

References

[1] Geoffrey Matthews and James Hearne. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175–184, 1991.

[2] R. S. Michalski and R. E. Stepp. Learning from observation: Conceptual clustering. *Machine Learning. An Artificial Intelligence Approach*, pages 331–363, 1983.

[3] Richard A. Reyment and K. G. Jöreskog. *Applied Factor Analysis in the Natural Sciences*. Cambridge University Press, New York, NY, 1993.

[4] W. G. Landis, R. A. Matthews, A. J. Markiewicz, and G. B. Matthews. Multivariate analysis of the impact of the turbine fuel jet-a using a standard aquatic microcosm toxicity test. *Environmental Sciences*, 2(3):113–130, 1993.

[5] W. G. Landis, R. A. Matthews, A. J. Markiewicz, and G. B. Matthews. Multivariate analysis of the impacts of the turbine fuel jp-4 in a microcosm test with implications for the evaluation of ecosystem dynamics and risk assessment. *Ecotoxicology*, 1994.

[6] Frieda B. Taub. Standardized aquatic microcosms. *Environm. Sci. Technol.*, 23:1064–1066, 1989.

[7] I. J. Good. An index of separateness of clusters and a permutation test for its significance. *J. statist. comp. Simul.*, 15:81–84, 1982.

Submitted to Applied Ecology

Nonmetric Conceptual Clustering in Ecology and Ecotoxicology

Geoffrey Matthews
Computer Science Department
Western Washington University

Robin Matthews
Huxley College of Environmental Studies
Western Washington University

Wayne Landis
Institute of Environmental Toxicology and Chemistry
Western Washington University

Send correspondence to: Geoffrey Matthews
Computer Science Department
Western Washington University, MS 9062
Bellingham, Washington, 98225, USA
matthews@cs.wwu.edu
(206) 650-3797

Running head: Nonmetric Conceptual Clustering

Abstract

Ecological studies and multispecies ecotoxicological tests are based on the examination of a variety of physical, chemical and biological data with the intent of finding patterns in their changing relationships over time. The data sets resulting from such studies are often noisy, incomplete, and difficult to envision. We have developed machine learning and visualization software to aid in the analysis, modelling, and understanding of such systems, and have applied it to the analysis of lake and stream field studies, and aquatic microcosm toxicological tests. The software is based on nonmetric conceptual clustering, which attempts to analyze the data into clusters that are strongly associated with several measured parameters. We have found in many cases that this approach is superior to classical clustering algorithms, all of which rely on an n-dimensional metric (or similarity measure). In each case, our tools not only confirmed suspected ecological patterns, but also revealed aspects of the data that were unnoticed by ecologists using conventional statistical techniques. Machine learning tools should, accordingly, become a standard part of the ecologist's armamentarium.

Introduction

Machine learning has fallen on hard times. Edward Feigenbaum, in a plenary talk at the recent IEEE conference on AI applications called it a "big disappointment."

Understanding ecosystems requires the solution of novel data analysis problems. Typically, dozens to hundreds of species, as well as many physical and chemical parameters, are sampled in natural and artificial systems. These parameters not only change over time, but sampling limitations necessitate acquiring only a few samples, resulting in shallow data matrices with many dimensions, but few points. The essential task of computational assistance, then, is to reduce the dimensionality and aid in the interpretation of these data sets. Nonmetric conceptual clustering was designed for these kinds of data (Matthews and Hearne, 1991). It simultaneously reduces both the complexity and the dimensionality of the set of data points. The complexity is reduced by grouping the points into clusters. The dimensionality of the data is reduced by selecting only parameters that fit well with the generated clusters. Random or noisy parameters are ignored. The ability to evaluate a model of the data simultaneously on several different fitness criteria gives nonmetric conceptual clustering its strength.

We have applied nonmetric clustering successfully in multispecies field and laboratory studies, and in each case we have not only confirmed the presence of suspected patterns, but also discovered aspects of the data that were unnoticed by ecologists (Landis et al., 1993; Matthews et al., 1991a; Matthews et al., 1991b). In addition, these patterns were usually overlooked by conventional statistical techniques. In this sense, the software has stepped beyond the role of traditional expert systems, which merely mimic human expertise, and into the role of a machine learning system: a computer system that can learn things about the data that a human cannot. Such systems bring a kind of power to human investigators, expertise that is beyond their own ability but which can form part of a valuable partnership.

We present here a summary of the nonmetric conceptual clustering approach, some results stemming from applications in ecology and ecotoxocology, and our attempts to extend the applicability of the nonmetric clustering paradigm to system dynamics.

Nonmetric Clustering

Nonmetric clustering is similar to conceptual clustering in that the clustering is designed, not only to fit the data, but also to create a *simple* and *conceptual*

description of the data (Michalski and Stepp, 1983; Fisher and Langley, 1986). The goal of nonmetric clustering is a partition of the data into disjoint and exhaustive subsets (the clusters) such that most of the points can be described by simple conjunctive descriptions involving some of the original parameters (canonical dimensions, *i.e.* without rotation, *etc*). For example, if a large number of the points (cluster A), in dimensions x, y, and z, had "medium", "small", and "large" values, respectively, and another large number of points (cluster B), had "large", "medium", and "medium" values on these same dimensions, then the points could be described by the two concepts:

$$\begin{aligned} \text{Cluster A:} &\Leftrightarrow (x = \text{medium}) \wedge (y = \text{small}) \wedge (z = \text{large}) \\ \text{Cluster B:} &\Leftrightarrow (x = \text{large}) \wedge (y = \text{medium}) \wedge (z = \text{medium}) \end{aligned}$$

If these two sets of points comprised nearly all of the original data, then the clustering would be complete. There may be other dimensions in the original data set, other than x, y, and z, but these dimensions would be regarded as irrelevant to the above clustering if x, y, and z sufficed.

To this end, the nonmetric clustering algorithm performs a (nonexhaustive) search through the space of all clusterings (partitions) of the data, and all divisions of the parameters into categories (*e.g.*, "small", "medium", and "large"), and all subsets of parameters. The search terminates when it finds a clustering, parameter subset, and categorical division, such that the fit to the data cannot be improved. Naturally, the space of partitions and divisions is too large to be searched exhaustively. Accordingly, a hill-climbing algorithm is employed, starting from a random partition and quantile divisions of the dimensions. The search is then repeated, starting from a different random initialization, to avoid local maxima. In our experience with both synthetic and real data, about ten repetitions are sufficient to avoid local maxima. The algorithm has been implemented in a computer program called Riffle, together with a graphical front end for viewing the results.

Nonmetric clustering has the following advantages over some conventional clustering methodologies:

- It works well with incomplete data, where several points may have missing values for a few dimensions.
- It works equally well with categorical, ordinal, and numeric dimensions.
- It does not require *ad hoc* modifications of the numeric dimensions, such as normalizing the variance.

- It does not rely on a metric, such as the Euclidean metric, which will combine parameters by sums of squares or other mathematical methods.

- It has the ability to ignore noisy parameters, *i.e.* parameters with a large variance but random with respect to the overall pattern. Size of the variance is not taken into account since all values on all dimensions are merely regarded as small, medium, or large.

The clustering itself is informative, but Riffle actually provides the user with more than a traditional clustering algorithm. It also reports a list of the parameters that have a strong association with the clusters is also revealing. This list, which is a subset of all of the parameters, records only those that are important or significant in relation to the patterns in the data. Parameters that vary randomly are automatically be excluded from the list.

There are a number of synthetic data sets on which Riffle can outperform traditional clustering algorithms (Matthews and Hearne, 1991). However, the most amazing successes with Riffle have been in the analysis of ecological and ecotoxicological data sets, which we describe in the following sections.

Aquatic Ecology

In both lake and stream studies, Riffle has succeeded in obtaining intuitively meaningful clusters. In a one-year study of benthic macroinvertebrates in a small stream, Riffle grouped the samples exactly as a human expert would have done, one group consisting of "clean" water samples (mayflies, stoneflies, etc.), and another group consisting of "dirty" water samples (flies, oligochaetes, *etc.*) (Matthews et al., 1991a). Several rare species were found to have high association with these clusters, and thus were reported by Riffle as important to the overall pattern. But these same species had been overlooked as important indicator species because of their rarity. The samples were collected over an entire season, and included both low-density and high-density samples as the benthos matured over the summer. Standard clustering techniques were confounded by this seasonal variance and grouped the samples into "early" and "late" samples, without regard to the fine structure of the populations.

In a multi-year study of physical/chemical parameters in a large monomictic lake, Riffle accurately clustered samples according to season into summer epilimnion and hypolimnion, as well as winter mixed water samples (Matthews et al., 1991b). In a result surprising to the investigators, it also identified a fourth class of samples. Upon reinvestigating, we noticed that this class had actually been

sampled from within the metalimnion—an unforeseen accident of the experimental design. Further clustering by Riffle of the biological data showed a strong correlation with the clustering of the physical/chemical parameters. Conventional clustering algorithms were not able to identify these patterns.

Ecotoxicology

Riffle has also been successful in analyzing data from synthetic microcosms, in particular, the Standardized Aquatic Microcosm, or SAM (Taub, 1989). In the SAM, twenty-four jars of water are prepared identically with several species of algae, *Daphnia*, and other biota. The jars are divided into four treatment groups, normally a control and three increasingly toxic doses. The jars are monitored closely for two months and population counts for all species, as well as physical/chemical parameters, are recorded every few days. Nonmetric clustering by Riffle can often pick out the four treatment groups from the biological data alone.

Under controlled situations, such as the SAM, nonmetric clustering can form the basis of a confirmatory statistical test, which we have termed nonmetric clustering and association analysis (NCAA). In this case, the known treatment groups form one categorical label, and the cluster numbers form another. (Sometimes, although by no means always, the treatment groups form an ordinal, and not merely categorical variable.) The association between treatment group and cluster number forms the basis of a confirmatory statistic: under the null hypothesis, there would be no association. Any contingency table test, such as the χ^2 test, can then be used to obtain a confidence level.

Nonmetric clustering consistently reveals aspects of the SAM microcosms that are hidden from other tests. Since Riffle reduces the dimensionality of the SAM by indicating which species are important on which days of the test, it gives the practitioner a good handle on how the populations respond to the toxin. Quite often one species will be important early in the test, of little importance during the middle period, and then important again later. We have also noticed "chaotic" trends in the evolution of the SAM. For instance, in at least two of the experiments, the treated groups diverged significantly from the control group, and then, by about the end of the first month, "recovered" to a state indistinguishable from the control group. However, during the second month, the treatment groups again diverged, in a dose-response fashion. This indicates that, during the putative recovery period, the systems were nonetheless quite different, and were able to diverge later. This is symptomatic of chaotic systems, where imperceptible differences in initial conditions can lead to radically different behavior subsequently.

Other Applications

Riffle is currently being applied to a wide variety of data analysis problems. We are currently beginning an investigation into the toxicity of refinery effluents, using measurements required by the National Pollution Discharge Elimination System (NPDES). Also, in cooperation with Dr. Anne Fairbrother of the U.S.E.P.A., Corvallis, we are applying Riffle to studies of biomarkers of toxicological impacts on mice and birds. Other researchers have applied Riffle to medical diagnosis problems.

Future Directions: Temporal Dynamics

As well as Riffle works in analyzing data, it is essentially static. Many of the effects seen in ecological data analysis are dynamic—an effect may be simply a time delay, for example. Further, oscillations, such as those in the predator-prey models, can be expected, as well as chaotic dynamics. We are beginning to apply the lessons learned from nonmetric clustering to the analysis of dynamic multivariate data. Some of our approaches are outlined below.

Discrete curvature and torsion: The path of an ecosystem through n-dimensional space over time can be viewed as a parameterized curve. Using analogies of the Frenet formulas (O'Neill, 1966, pp. 56–66), discrete analogues of the fundamental vectors, velocity, curvature, torsion *etc.*, can be defined and used to characterize the evolution of the system.

Nonmetric clustering strain: The key idea behind nonmetric clustering strain is to measure the change in nonmetric clustering from one time slice to the next. By examining how nonmetric clusters of the points change over time, measures of the size and direction of the change can be obtained.

Conceptual shift: When performing conceptual clustering the important parameters usually change over time. Thus, not only do the points change their relationships, but the conceptual descriptions of the points can use a different vocabulary at different times. The measure of how the "best" description changes over time gives us another handle on understanding dynamic behavior.

Visualization: We are also investigating graphical visualization of the evolution of systems in n-dimensional phase space over time. The curvature, torsion, clustering shift and conceptual shift can all be visualized with interactive computer graphics. Projection pursuit and grand tour algorithms can be used

to maximize the visibility of desired quantities (Asimov, 1985; Huber, 1985). Critical points, at which the behavior of the systems becomes "interesting," can then often be found by inspection.

Conclusions

Our program attempts to understand multivariate data on its own terms. To this end, we have built and applied nonmetric clustering and visualization tools that reduce the dimensionality and complexity of multispecies systems to a manageable size. Other attempts have been made to understand ecosystems in terms of multivariate response, but the responses were usually measured using n-dimensional metrics (Johnson, 1988; Kersting, 1988). We have seen repeatedly that metric approaches suffer from a large number of drawbacks when dealing with ecological data. The approach recommended here is free from any metric (or similarity measure) and its problems.

Recently, the U.S. Environmental Protection Agency has instituted a policy that calls for the cancellation of multispecies toxicity tests because data analysis has proven too difficult or inconclusive (Fisher, 1992). We believe that the problem is not with the multispecies tests, which are carefully designed to be more realistic than classic, single-species tests, but rather with the poor quality of the data analysis tools that are applied to the results of these tests. So far as we know, we are the only group in the United States applying the methodologies of machine learning to multivariate ecological and ecotoxicological studies, and we are seeing results that greatly enhance our understanding of the systems and their dynamics. Interest in our techniques at national toxicological conferences is always high, and we are convinced that the machine learning paradigm will revolutionize ecology and ecotoxicology in the near future.

References

Asimov, D. (1985). The grand tour. *SIAM Journal of Scientific and Statistical Computing*, 6:128–143.

Fisher, D. and Langley, P. (1986). Conceptual clustering and its relation to numerical taxonomy. In Gale, W. A., editor, *Artificial Intelligence and Statistics*, pages 77–116. Addison Wesley.

Fisher, L. J. (1992). Decisions of the ecological, fate, and effects task force. *U.S. E.P.A. Memo, October 29.*

Huber, P. J. (1985). Projection pursuit. *Annals of Statistics*, 13:435–475.

Johnson, A. R. (1988). Evaluating ecosystem response to toxicant stress: a state space approach. In Adams, W. J., Chapman, G. A., and Landis, W. G., editors, *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP 971*, pages 275–285. American Society for Testing and Materials, Philadelphia.

Kersting, K. (1988). Normalized ecosysterm strain in micro-ecosystems using different sets of state variables. *Verhandlungen der Internationalen Vereinigung fuer Theoretische und Angwandte Limnologie*, 23:1641–1646.

Landis, W. G., Matthews, R. A., Markiewicz, A. J., and Matthews, G. B. (1993). Multivariate analysis of the impact of the turbine fuel jet-a using a standard aquatic microcosm toxicity test. *Environmental Sciences*, 2(3):113–130.

Matthews, G. and Hearne, J. (1991). Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175–184.

Matthews, G. B., Matthews, R. A., and Hachmöller, B. (1991a). Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(11):2184–2190.

Matthews, R. A., Matthews, G. B., and Ehinger, W. J. (1991b). Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modelling*, 53:167–187.

Michalski, R. S. and Stepp, R. E. (1983). Learning from observation: Conceptual clustering. *Machine Learning, An Artificial Intelligence Approach*, pages 331–363.

O'Neill, B. (1966). *Elementary Differential Geometry*. Academic Press, New York.

Taub, F. B. (1989). Standardized aquatic microcosms. *Environm. Sci. Technol.*, 23:1064–1066.

Application of Multivariate Techniques to Endpoint Determination, Selection and Evaluation in Ecological Risk Assessment

Wayne G. Landis[1], Geoffrey B. Matthews[2], Robin A. Matthews[1] and Anne Sergeant[3]

[1]Institute of Environmental Toxicology and Chemistry, Huxley College, Western Washington University, Bellingham, WA 98225

[2]Department of Computer Science, Western Washington University, Bellingham, WA 98225

[3]Anne Sergeant, ORD, U.S. EPA, 401 M Street, S.W. (RD-689), Washington DC 20460

Running Title: Multivariate Risk Assessment

Corresponding Author: Wayne G. Landis, IETC, Huxley College, Western Washington University, Bellingham, WA 98225-9079, 206-647-6136

Abstract: Ecological risk assessment has evolved so that the interaction among the components is now an implicit assumption. Unlike single species based risk assessments, it is often crucial in environmental or ecological risk assessments to be able to describe a system with many interacting components. In addition, some quantifiable description of how different biological communities are upon the addition of a toxicant or some other stressor is required to adequately describe risk at the ecosystem level. Three methods have been applied at the ecosystem level, the mean strain measurement used by K. Kersting, the state space analysis pioneered by A.R. Johnson, and the nonmetric clustering developed by G. Matthews for ecological datasets and for analysis of Standardized Aquatic Microcosm data. Each method has direct application to the description of an effected ecosystem without reliance upon a single and specific and perhaps misleading endpoint. Each also can assign distance or probability measures in order to compare the control to treatment groups. Nonmetric clustering (NMC) has the advantage of not attempting to combine different types of scales or metrics during the multivariate analysis and is robust against interference by random variables. Application of these methodologies into an ecological risk assessment should have the benefit of combining large interactive datasets into distinct measures to be used as a measure of risk and as a test of the prediction of risk. The primary impact of these methods may be in the selection and interpretation of assessment and measurement endpoints.

Much recent debate in toxicological studies has focused on appropriate endpoints for tests. Nonmetric clustering and other multivariate techniques should aid in the selection of these endpoints in ways meaningful at the ecosystem level. We suggest that the search for assessment and measurement endpoints be left to the appropriate multivariate computation algorithms in the case of multispecies situations. Application of these methods in the verification, validation process of risk assessment will prove to check the selection of endpoints during modeling exercises and to improve the presentation of assessment criteria.

Key Words: Risk assessment, multivariate statistics, nonmetric clustering, measurement and assessment endpoints, artificial intelligence.

Ecological Risk Assessment Defined

Ecological risk assessment is essentially the art of extrapolating from relatively straight-forward information on how toxic a compound is to specific organisms to how complex assemblages of organisms will respond to the toxin in their natural environment. The traditional approach to ecological risk assessment was developed by the National Academy of Science (NAS) using a human health effects paradigm. The NAS model is described in detail in *Risk Assessment in the Federal Government: Managing the Process* (1), also known as the "red book." The NAS approach uses a four-point approach:

a) The initial **hazard identification**, which determines whether a chemical is capable of causing adverse health effects. This conclusion is based on laboratory animal studies and, where available, human data;

b) The **dose-response assessment**, which characterizes the relationship between the chemical dose and the incidence of adverse health effects in the exposed population;

c) The **exposure assessment**, which measures or estimates the intensity, frequency, and duration of human exposure to a chemical, or estimates hypothetical exposure; and

d) The **risk characterization**, which combined the dose-response and exposure assessments. This final step evaluates the uncertainties in the previous analyses and provides an estimate of the likelihood of adverse effects under the stated conditions.

The NAS paradigm was developed to assess the risks of chemicals to human health, and while many of its principles can be implemented directly in ecological risk assessment, it falls short when applied to non-chemical stressors or interdependent organisms. Furthermore, it does not even begin to address the links between organisms and their environment. Hazard identifications are complicated by the many metabolic and degradation pathways available in the environment. Changes in these pathways can occur naturally, as a result of spatial and temporal changes in species assemblages, but can also be induced as a result of the introduction of a xenobiotic. Exposure assessments are complicated by the extraordinary array of species present at the exposure sites. The species composition also changes as a result of natural forces (seasonality, stochastic extinctions, migrations, etc.) or the introduction of a xenobiotic. Because of this, ecological risk assessment must be recognized as being fundamentally different from human health risk assessments (2).

Ecological Risk Assessment Models - Review of the USEPA Framework

Many of the difficulties in applying the traditional risk assessment paradigm to ecosystems have been addressed in the recent formulation of a *Framework for Ecological Risk Assessment* (3) (Figure 1). Among the novel features of this framework is the integration of exposure and hazard assessment to reflect the interactions that occur in ecological systems. Also innovative is the inclusion of a Data Acquisition, Verification and Monitoring process within the framework. The key however, is the selection

of assessment and measurement endpoints to make the assignment of risk representative of the system under protection.

The USEPA Framework includes three steps: problem formulation, analysis, and risk characterization.

Problem formulation is the process that evaluates the characteristics of the stress-inducing agent (e.g., toxin). It also identifies the ecosystem that may be at risk, and identifies possible ecological effects. This information is used to select the ecosystem components or attributes of concern (the assessment endpoints) and to determine the best ways to describe this component or attribute (measurement endpoints). Finally, the assessor prepares a conceptual model that describes the ways in which the stressor could interact with the ecosystem and the likely effects of such an interaction. Problem formulation is not specifically discussed in the NAS paradigm, but in current practice these issues are addressed during planning.

The **analysis** phase contains two components: **characterization of exposure** and **characterization of ecological effects**. The exposure characterization determines stressor distribution, characterizes receptors, and quantifies stressor release, migration, and fate. The effects characterization evaluates effects data and response data such as stressor-response analysis (akin to the dose-response assessment described above), the relationship between endpoints, and evidence of causality. This phase is analogous to the hazard identification, dose-response and exposure assessment components of the NAS paradigm.

The **risk characterization** component differs little from its counterpart in the NAS paradigm. It tests the hypotheses developed in the conceptual model described in Problem Formulation by synthesizing information about the stressor and receptor from various sources and describing the supporting evidence for (and uncertainty associated with) conclusions. It also provides some indication of the likelihood of effects occurring and describes the ecological significance of any predicted risk.

Endpoint Selection-Ecological Risk Assessment

Endpoints (assessment and measurement) are the keystones of an ecological risk assessment as every other parameter in the process is predicated upon these terms. An assessment endpoint must be something specific and quantifiable such as "maintenance of sport fish populations" or "desertification" or "eutrophication." Values such as "ecosystem health" have little meaning (2) and cannot be easily described. Sometimes it is not possible to examine the assessment endpoint directly--for example, one cannot collect bald eagle livers and analyze them for enzyme induction. In this case, measurement endpoints are used to describe the organism or entity of concern. Continuing with the bald eagle example, one may wish to examine contaminant concentrations in the eagles' food and compare them to laboratory dose-response data, observe their feeding habits and construct exposure scenarios, and review liver-enzyme data from other eagles (in captivity or found dead) or other birds of prey to arrive at

conclusions about enzyme induction in local eagles. In the ecosystem sense, measures of species number, abundance or energy flow would be analogous.

The USEPA Framework recommends that assessment endpoint selection consider 1) ecological relevance, 2) policy goals and societal values, and 3) susceptibility to the stressor. To ensure that ecological relevance is addressed, one must have some *a priori* knowledge of the ecosystem of interest and the relationships between its components. Science must not take a back seat to policy and societal values, but communication between the risk assessor and risk manager is critical to ensure scientific integrity and satisfy policy needs. Finally, the strongest assessment endpoints are both affected by the stressor and sensitive to a specific type of effect caused by that stressor.

Measurement endpoints should be selected on the basis of how well they represent assessment endpoints. Practicality and consistency with exposure scenarios often determine the initial range of possibilities. Measurement endpoints must be correlated with or useful for inferring changes in assessment endpoints (4). To the extent possible, they should be selected for appropriate diagnostic ability, signal-to-noise ratio, sensitivity, and response time. Ideally, measurement endpoints also provide information about indirect effects such as toxicity to an organism upon which the species of interest preys or nutrient cycle inhibition reducing survivorship of fingerlings.

An ecological risk assessment is only as good as the data upon which it is based. Thus, data acquisition is an integral part of the risk assessment process. Endpoints can and generally should change with time. At any stage in ecological risk assessment, new data may reveal that a particular endpoint should be added or removed, or that it no longer provides relevant information. For example, tree seedling success may be an important measure in managed ecosystems or when bare or disturbed soil is being colonized, but it provides little information about old-growth forests. Similarly, a measure of biomass in an aquatic system may provide a good indication of overall productivity, but it probably will not contain enough information to determine whether a balanced assemblage of functional groups (shredders, filter-feeders etc.) exists. Preliminary data needs should be outlined during the Problem Formulation and refined as needed during the rest of the risk assessment process. For example, the assessor may discover that the assessment endpoint initially selected is affected less by the stressor being evaluated than by other causes, such as widespread habitat loss or overfishing--this may require selection of another assessment endpoint. Similarly, as the assessment progresses, it may become evident that additional measurement endpoints are needed. Increasingly, the use of multivariate data analysis is being called upon to assist in identifying appropriate endpoints for ecological risk assessments.

Importance of Multivariate Data in Ecological Risk Assessments

One important feature of ecological risk assessments is that they generally must rely on multivariate data to identify natural and toxicant-induced patterns. This is a result of the multidimensional nature of ecosystems; the Hutchinsonian idea of organisms and populations residing in a n-dimensional

hypervolume is the basis of current niche theory (5). The n-dimensional hypervolume is the ecosystem with all its components as perceived by the population. The variability of these parameters over time as well is used to account for the variety of species within the ecosystem system (6,7,8). Applications of resource competition models have been proposed for evaluating even single-species toxicant effects (9). Therefore, in order to begin to describe an ecosystem's response to perturbation, we must recognize the system's multidimensional nature.

Our essential goal in multivariate data analysis is to identify ecologically relevant patterns in the data set. This is true regardless of whether our ultimate goal is to develop an ecological risk assessment or to evaluate naturally occurring changes in the ecosystem. However, until recently, the data reduction tools available to aid our analyses have consisted primarily of simple graphs (lots of them), simple statistical tests done repeatedly to accommodate all of the measured parameters, and a few truly multivariate statistical tests that generated useful but esoteric results. For example, analysis of variance (ANOVA) is the classical method to examine single variable differences from control groups or reference sites. However, in multivariate data, there are problems with Type II errors. Furthermore, it is difficult to display and assimilate the many ANOVA results that are generated from a multivariate data set. Conquest and Taub (10) developed a method to overcome some of these problems by generating intervals of non-significant difference for a single variable measured repeatedly over time. This method corrects for the likelihood of a Type II error and produces a visual display of significant vs. nonsignificant differences that is easily graphed. The major drawback to this method is that it only portrays changes in single variables over time.

Multivariate methods have proved promising as a method of incorporating all of the dimensions of an ecosystem. One of the first to be used in toxicology was the calculation of ecosystem strain developed by Kersting (11,12,13,14) for relatively simple (three species) microcosms. At about the same time, Johnson (15,16) developed a multivariate clustering algorithm to map the n-dimensional coordinates of an ecosystem and used the distance between these systems as a measure of divergence from the control. Both of these methods have the advantage of examining the multispecies test systems as a whole and can track such process as succession, recovery and the deviation of a system due to an anthropogenic input. Their major disadvantage, which is also a disadvantage with most conventional multivariate statistical techniques, is that all of the data are incorporated without regard to the metric (unit of measurement) or relative value of a variable toward identifying patterns in the data set ("noisy" or random data are included along with the rest). It can be difficult to reconcile variables such as pH with a 0-14 metric to the numbers of bacterial cells per mL, where low numbers are in the 10^6 range. Along the same lines, data that vary randomly and have large metrics may overwhelm the statistical computations and mask the importance of highly correlated variables with small metrics.

Ideally, multivariate statistical tests used for evaluating complex data sets, whether the goal is to develop an ecological risk assessment or not, will have the following characteristics:

a) It will not combine counts from dissimilar taxa by means of sums of squares, or other ad hoc mathematical techniques, as in the Euclidean and cosine distance measures;

b) It will not require transformations of the data, such as normalizing the variance;

c) It will work without modification on incomplete data sets;

d) It will work without further assumptions on different data types (e.g., species counts or presence/absence data);

e) The Significance of a taxon to the analysis will not be dependent on the absolute size importance with common taxa, and taxa with a large, random variance will not automatically be selected to the exclusion of others;

f) It will provide an integral measure of "how good" the clustering is, i.e. whether the data set differs from a random collection of points; and

g) It will, if appropriate, identify a subset of the taxa that serve as reliable indicators of the physical environment.

Although we have now defined the ideal characteristics of a multivariate system, none is of course perfect. However, a method borrowed from the Artificial Intelligence (AI) tradition meets a large proportion of the above design criteria.

Nonmetric Clustering and Association Analysis

Unlike the more conventional multivariate statistics, nonmetric clustering is an outgrowth of artificial intelligence and a tradition of conceptual clustering. In this approach, an accurate description of the data is only part of the goal of the statistical analysis technique. Equally important is the intuitive clarity of the resulting statistics. For example, a linear discriminant function to distinguish between groups might be a complex function of dozens of variables, combined with delicately balanced factors. While the accuracy of the discriminant may be quite good, use of the discriminant for evaluation purposes is limited because humans cannot perceive hyperplanes in highly dimensional space. By contrast, conceptual clustering attempts to distinguish groups using as few variables as possible, and by making simple use of each one. Rather than combining variables in a linear function, for example, conjunctions of elementary "yes-no" questions could be combined: *species A greater than 5, species B less than 2, and species C between 10 and 20.* Numerous examples throughout the artificial intelligence literature have proven that this type of *conceptual* statistical analysis of the data provides much more useful insight into the patterns in the data, and is often more accurate and robust. Delicate linear discriminants, and other traditional techniques, chronically suffer from overfitting, particularly in highly dimensioned spaces. Conceptual statistical analysis attempts to fit the data, but not at the expense of a simple, intuitive result.

Applications of Nonmetric Clustering and Association Analysis

A detailed description of our multivariate methods, including nonmetric clustering and association analysis is in Appendix A. As examples of the usefulness of multivariate methods in general, and nonmetric clustering in particular, we will use examples of field evaluations and toxicity tests conducted over the last 3 years. Insights into the utility of these methods, the dynamics of even straightforward microcosm systems, and the importance of measurement variables have been the results of these studies.

Field Studies

Before we can determine whether a toxin has affected a group of organisms or the dynamics of an ecological community, we must first determine what types of changes would occur that are independent of the toxin. In field situations, this is usually attempted by using a reference site, monitoring the changes that occur at that site, and comparing this with the changes that occur in organisms at the "treatment" site.

However, one of the most difficult analytical challenges in ecology is to identify patterns of change in large ecological data sets. Often these data are not linear, they rarely conform to parametric assumptions, they have incommensurable units (e.g., length, concentration, frequency, etc.), and they are incomplete (due to both sample loss and sampling design whereby different parameters are collected at different frequencies). These difficulties exist regardless of whether there are toxins present; the only difference is that with the presence of a toxin, we must try to separate the response to the toxin from the other changes that occur at the site(s).

We have compared several types of multivariate techniques to evaluate two types of ecological data, a limnological data set that included spatial and temporal changes in water chemistry and phytoplankton populations, and a stream data set that included spatial (longitudinal) and temporal changes in benthic macroinvertebrate species assemblages (17,18) . Our objective was to see whether the multivariate tests could identify obvious patterns involving the influences of stratification in the lake and the effects of substrate and water quality changes on stream macroinvertebrates. We used principal components analysis, hierarchical clustering (k-means with squared Euclidean or cosine of vectors distance measures), correspondence analysis, and nonmetric clustering to look for patterns in the data.

In both studies, nonmetric clustering outperformed the metric tests, although both principal components analysis and correspondence analysis yielded some additional insight on large-scaled patterns that was not provided by the nonmetric clustering results. However, nonmetric clustering provided information without the use of inappropriate assumptions, data transformations, or other data set manipulations that usually accompany the use of multivariate metric statistics. The success of these studies and techniques lead to the detailed examiniation of community dynamics in a series of two multispecies toxicity tests.

Multispecies Toxicity Testing

The multivariate methods described above have recently been used to examine a series of multispecies toxicity tests. Described below are the data analyses from two recently published tests using methodology derived from the Standardized Aquatic Microcosm (SAM) (ASTM E1366-91). The 64-day SAM-protocol previously has been described (19,20,21,22,23). Briefly, the microcosms were prepared by the introduction of ten algal, four invertebrate, and one bacterial species into 3L of sterile defined medium.

In the first example (24), the riot control material 1,4-dibenz oxazepine (CR) was degraded using the patented organism *Alcaligenes denitrificans denitrificans* CR-1 (*A. denitrificans* CR-1). *A. denitrificans* CR-1 was obtained using a natural inoculum set in an environment containing the microcosm medium T82MV containing the toxicant CR. After demonstrating the organisms ability to degrade the toxicant CR, a microcosm experiment was set up to investigate the ability of the microorganisms to degrade CR in an environment resembling a typical freshwater environment. Toxicity tests of the riot control material demonstrated that although *A. denitrificans* CR-1 eliminated the toxicity of a CR solution towards algae, toxicity did remain to *Daphnia magna.*

The SAM experiment was set up with a control group without the toxicant or *A. denitrificans* CR-1, a second group with only CR, a third group with only *A. denitrificans* CR-1, and the fourth group containing both the toxicant CR and the bacterium *A. denitrificans* CR-1. Conventional analysis demonstrated that the major impact was the increase in algal populations since both CR and the degradative products of the toxicant both inhibited the growth of the major herbivore, *D. magna.* The control group and the microcosms inoculated initially with *A. denitrificans* CR-1 were not distinguishable using conventional analysis.

As a first test of the use of multivariate analysis in the interpretation of multispecies toxicity tests, the data set used to analyze the CR microcosm experiment were presented in a blind fashion for analysis. Neither the purpose of the experiment or the experimental set up was provided for the analysis. Nonmetric clustering was used to rank variables in terms of contribution and to set clusters. Surprisingly, the analysis resulted in only two clusters being recognized, Control and *A. denitrificans* CR-1 treatments, and the CR and CR plus *A. denitrificans* CR-1 treatments. Variables important in assigning clusters were D. magna, Ankistrodesmus, Scenedesmus and NO_2. Obviously, the inclusion of the principal algal species in these experiments and the daphnia was not a surprise, but NO_2 had not been demonstrated as a significant factor in previous analysis. However, the species *A. denitrificans denitrificans* is classified for its denitrification ability (25).

The second major application of nonmetric clustering to the analysis of SAM data has been the investigation of the impact of the water soluble fraction (WSF) of the fuel Jet-A (26). Four treatment groups, control, 1, 5 and 15 percent WSF were used.

All of the multivariate tests (cosine distance, vector distance and nonmetric clustering) agree that a significant difference between treatment groups was observed through day 25. From day 28 to day 39, the effect diminished until there were no significant effects observable. However, significant effects were again observable from day 46 through day 56, after which they again disappeared for days 60 and 63.

In Figure 2, the average cosine distances within the control group and between the control group and each of the three treatment groups are plotted on a log scale. The initial, strong effect, from day 11 to day 25, is easily seen as a large distance from the treatment 1 (control) and treatment 2, together, to both treatment groups 3 and 4, initially, but then treatment 3 moves closer to the control. The period of no significant difference, from day 35 to day 46, is also clear. During the second period of significant difference, from day 49 to 59, a perfect dose-response for all three treatments is seen, with higher doses becoming more distant from the control. This dose-response relationship is consistently maintained over a period of eleven days, for four sampling dates, days 49, 53, 56, and 59. In general, a dose-response relationship like this was not observed earlier, although the magnitude of the distance was considerably greater.

Also of interest are the variables that best described the clusters and the stability of the importance of the variables during the course of the experiment. Table 1 lists the variables determined to be important in determining the clusters by importance for each sampling day as determined by nonmetric clustering. In general, the number of variables that were important was larger during the start of the test and lower at the end. In addition, a great deal of variability in rankings is apparent during the course of the SAM. The number of sampling dates when a variable was deemed important in cluster formation is listed in Table 2. Ankistrodesmus was the most consistent of the variables, being ranked in 12 out of the 16 sampling dates. Medium daphnia was also ranked often. However, variables like Ostracod and Philodina did not become important until later in the experiment.

The repeated oscillation of the dosed replicates compared to the controls were accounted for in two basic ways:

a reflection of the functioning of the community best described by parameters not directly sampled by the SAM protocol; or,

a repeated fluctuation in community structure initiated by the initial stress and that is visible as an undampened movement in the systems.

Until more data can be obtained, the cause-effect of the second oscillation can not be determined. However, the use of multivariate analysis detected an unexpected result, one providing a new insight into the dynamics of even the relatively simple laboratory microcosm.

Synthesis

Several other researchers have attempted to employ multivariate methods to the description of ecosystems and the impacts of chemical stressors. Perhaps the best developed approaches have been those of K. Kersting and A.R. Johnson.

Multivariate Descriptions of Microcosm Systems

Normalized Ecosystem Strain (NES) was developed by Kersting (11,13) as a means of describing the impacts of several materials to the three compartment microecosystems containing an autotrophic, herbivore and decomposer subsystems. These variables in the unperturbed control systems are used to calculate the normal operating range (NOR) of the microecosystem. The NOR is the 95 per cent confidence ellipsoid of the unperturbed state of a system. The center of the NOR is defined as the reference point for the calculation of the NES. The NES is calculated as the quotient of the Euclidean distance from a state to the reference state divided by the distance from the reference state to the 95 percent confidence (also called tolerance) ellipsoid, along the vector that connects the reference state to the newly defined state. A value of 1 or less indicates that the new state is within the 95 percent confidence ellipsoid, values greater than 1 indicate that the system is outside this confidence region.

Originally limited to ellipsoids, the use of Mahalonobis distances allows the use of more variables as the confidence ellipsoid can be transformed to a confidence or tolerance hypersphere. These ideas were examined using the microecosytem test method developed by Kersting for the examination of multispecies systems. In tests using a relatively straightforward multicompartment microcosm the sensitivity and strengths of this methods were observed. The sensitivity of the NES increased sensitivity as the number of variables used to describe the system increased (13). Another interesting observation was the increasing distance from the normal space of the system after a perturbation as measured by NES as time increased. This increasing distance indicates that the perturbed system is drifting from its original state. Kersting hypothesized that the system may even shift to a different equilibrium state or domain and that the system would remain there even after the release of the stressor.

Apparently as an independent development, A.R. Johnson (15) proposed the idea of using a multivariate approach to the analysis of multispecies toxicity tests. This state space analysis is based upon the common representation of complex and dynamic systems as an n-dimensional vector. In other words, the system is described at a specific moment in time as a representation of the values of the measurement variables in an n-dimensional space. A vector can be assigned to describe the motion of the system through this n-dimensional space to represent successional changes, evolutionary events, or anthropogenic stressors. The direction and position information form the trajectory of the state space and this can be plotted over time.

In the n-dimensional hypervolume that describes the placement and trajectory of the ecosystem it is possible to compare the positions of systems at a specified time. This displacement can be measured by

literally computing the distance from the systems and this displacement vector can be regarded as the displacement of these systems in space. This displacement vectors can be easily calculated and compared. Using the data generated by Giddings (27) in a series of classic experiments comparing results of the impacts of synthetic oil on aquarium and small pond multispecies systems, Johnson was able to plot dose response curves using the mean separation of the replicate systems. These plots are very reminiscent of dose-response curves from typical acute and chronic toxicity tests.

As summarized by Johnson, the strengths of this methodology are the objectivity for quantifying the behavior of the stressed ecosystem and the power of this methodology to summarize large amounts of data. As with the work of Kersting, this methodology allows the investigator to examine the stability of the ecosystem and the eventual fate of the system relative to the control treatment.

Another important application proposed by Johnson (16) was the use of multivariate analysis to identify diagnostic variables that can be applied in the monitoring of ecosystems. Diagnostic variables, if reliable in differentiating anthropogenically stressed systems from control systems would be extremely valuable in monitoring for compliance and in determining clean up standards. The use of such variables is justified due to the fact that decisions often have to be made with incomplete datasets due to technical difficulties, cost, and a general lack of knowledge. Techniques proposed for the determination of these variables included linear regression, discriminant analysis and visual inspection of graphed data. Johnson conducted a cost-benefit analysis using an ecosystem model that demonstrated under the condition of that model, the benefits of diagnostic variables. In the Discussion, Johnson proposes simulation modeling to attempt to find generalized diagnostic variables that best describe the state space and trajectory of an ecosystem.

The major difficulty with the methods detailed above is the reliance on conventional metric statistics. Vector distances in an n-dimensional space including such disparate variables as pH, cells counts and nutrient concentrations are difficult to compare from one experiment to another. Another consideration is the fact that many of the variables may be compilations of others. Algal biomass is often calculated by using multiplying cell counts by an appropriate constant for each species. Species diversity and many indices of ecosystem health are similarly composited variables. As discussed in the pervious sections, the use of metric methods with nonmetric clustering may prove a useful combination.

Search for Relevant Assessment and Measurement Endpoints

The attempt by Johnson to derive diagnostic variables is an interesting approach. However, our current research indicates that identity of the variables that contribute the most to separating control treatment from dosed treatment groups change from sampling period to sampling period. The variables change in the SAM experiments, no doubt, in response to the successional trajectory of the system as nutrients become depleted. As nutrients become limiting and the ability of the system to exhibit large

differences in community structure become less, the metric measures do not exhibit the same magnitudes of separation. Nonmetric clustering does not seem to be as sensitive to these changes.

However, the search for diagnostic measures to indicate the displacement of an ecosystem may not be fruitless. Although the relative importance of the variables in the SAM experiments may change, there are often variables that are more critical during the earlier stages of the development of the microcosm and those that are more crucial in the latter stages. The variable Ostracods is generally more important in the latter half of the experimental series than in the latter stages. The crucial aspect is that the clustering algorithm is able to select ecosystem attributes that are the best in differentiating stressed versus non-stressed systems. Although expert judgment may be able to predict in some cases variables that could be considered important to measure, the clustering approach is rapid, consistent, and not biased.

Instead of defining Assessment Endpoints, it may be more practical to define an Assessment Baseline or hypervolume using variables that have been demonstrated to be important in past descriptions of these types of ecosystems. Defining the 95 percent confidence region may be a more accurate way of characterizing the problem than by using artificial constructs or individual assessment measurement endpoint combinations. Assignment of these confidence regions may also improve the quality and accuracy of environmental risk assessment. Another logical outcome is that these regions must be defined by the measurement endpoints (variables). Measurement endpoints are the means by which a system can be accurately placed and its trajectory defined in an n-dimensional coordinate system. Such a means of describing systems has already been proposed by Kersting. The confidence region used to calculate NES is static, but an accounting of the passage of such a system through the coordinate system should provide a region from which deviation can be measured. Comparing dosed treatment groups to a control group is essentially the corresponding exercise but using a control series of replicates instead of an *a priori* prediction to measure deviation from the Assessment Baseline hypervolumes.

Measurement endpoints are therefore operationally defined, in the context of this paper using a multivariate approach, as the variables the set the axes for the description of the system within the n-dimensional space. Data such as dose-response curves may play a part if they describe a relevant axes when used in a biomonitoring role. Dose response data, however, are not measurement endpoints by themselves, but are important in setting relevant system parameters. It is preferable to select measurement endpoints that are the lowest common denominator of the system that is capable of being measured. For example, pH is certainly the most direct measurement of hydrogen ion concentration available. Diversity and other indices of species number and community structure, however, are composites of species abundance data.

The Myth of Ecosystem Health and Measurement Indices

The use of indices such as diversity and the Index of Biological Integrity have the effect of collapsing the dimensions of the hypervolume in a relatively arbitrary fashion. Indices, since they are composited variables, are not true endpoints. The collapse of the dimensions that are composited to tends to eliminate crucial information, such as the variability and distribution of the organisms within a particular system. The mere presence of absence and the frequency of these events can be analyzed using techniques such as nonmetric clustering and preserves the nature of the dataset. A useful function was certainly served by the application of these methods, but the new methods of data analysis and compilation should serve to replace these approaches and preserve the underlying structure and dynamic nature of ecological systems.

Part of the attraction of using indices may result in the pervasive nature of the metaphor, ecosystem health. In a recent critical evaluation, Suter (2) dismissed ecosystem health as a misrepresentation of ecological science. Ecosystems are not organisms with the patterns of homeostasis determined by a central genetic core. Since ecosystems are not organismal in nature, health is a property that can not describe the state of such a system. The urge to represent such a state as health has lead to the compilation of variables with different metrics, characteristics and casual relationships. Suter suggests a better alternative would be to evaluate the array of ecosystem processes of interest, a process that is now possible given multivariate methods.

Future Developments

Modeling of ecosystems may play an even more important role as the ability to generate the Assessment Baseline hypervolumes increases. However, the critical aspect is that these models not only predict the outcomes of the species under protection or the fishery that must be preserved but also the values of the measurements that can be made in a field or laboratory situation. These predictions should also predict sampling variability and chaotic and stochastic variation. The development of such models would be a critical development in the formulation of risk assessment methodologies.

Development of such models should be made with the understanding that the probability of divergence from the control state or the Assessment Baseline hypervolume given enough time will be 1.00. Assessment goals should be defined with reasonable time periods.

A major difficulty in the exploitation of these methods is that the vector distances, and to some extent even the cosine distances are not transferable or comparable unless the variables measured are essentially the same with the same metrics. Systems with different descriptive parameters will by definition occupy a different volume of n-dimensional space, making comparisons difficult. Determining the relevant parameters to use a measurement endpoints *a priori* may be difficult if not impossible.

There are benefits that should evolve directly from the use of multivariate techniques. First, it should force the description of measurement and assessment endpoints in terms of acceptable variance in a

dynamic fashion with expected distributions or functionality. Probabilistic criteria will certainly evolve from these aspects.

As these criteria are developed, the recognition that ecosystems are unique in their basic nature and not amenable to descriptions that incorporate only one dimensionally with that dimension an arbitrary axis.

Finally, the use of multivariate techniques should enable the researcher and assessor the capability of using all of the data in the description of an ecosystem with the results presentable to a decision maker or risk manager. After all, it has proven feasible to portray the results of these analysis in terms of distance and probabilities.

Acknowledgments. This research is supported by United States Air Force Office of Scientific Research Grant No. AFOSR-91-0291 DEF. We would also like to thank G. Suter, S. Norton, J. Dulka and B. Weber for their thoughtful discussions and encouragement.

Appendix A. Multivariate Techniques

In the research described below, three multivariate significance tests were used. Two of them were based on the ratio of multivariate metric distances within treatment groups vs. between treatment groups. One of these is calculated using Euclidean distance and the other with cosine of vectors distance (28,29) (Figure 3). The third test used nonmetric clustering and association analysis (30). In the microcosm tests there were four treatment groups with six replicates, giving a total of 24. This example is used to illustrate the applications in the derivations that follow.

Treating a sample on a given day as a vector of values, $\vec{x} = \langle x_1, \ldots x_{17} \rangle$, with one value for each of the measured biotic parameters, allows multivariate distance functions to be computed. Euclidean distance between two sample points $\vec{x}$ and $\vec{y}$ is computed as

$$\sqrt{\sum_i (x_i - y_i)^2}$$

The cosine of the vector distance between the points $\vec{x}$ and $\vec{y}$ is computed as

$$1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum y_i^2}}$$

Subtracting the cosine from one yields a distance measure, rather than a similarity measure, with the measure increasing as the points get farther from each other.

The within-between ratio test used a complete matrix of point-to-point distance (either Euclidean or cosine) values. For each sampling date, one sample point $\vec{x}$ was obtained from each of six replicates in the four treatment groups, giving a 24 x 24 matrix of distances. After the distances were computed, the ratio of the average within group metric (*W*) to the average between group metric (*B*) was computed (*W/B*). If the points in a given treatment group are closer to each other, on average, than they are to points in a different treatment group, then this ratio will be small. The significance of the ratio is estimated with an approximate randomization test (31). This test is based on the fact that, under the null hypothesis, assignment of points to treatment groups is random, the treatment having no effect. The test, accordingly, randomly assigns each of the replicate points to groups, and recomputes the *W/B* ratio, a large number of times (500 in our tests). If the null hypothesis is false, this randomly derived ratio will (probably) be larger than the *W/B* ratio obtained from the actual treatment groups. By taking a large number of random reassignments, a valid estimate of the probability under the null hypothesis is obtained as $(n+1)/(500+1)$, where n is the number of times a ratio less than or equal to the actual ratio was obtained (31).

In the clustering association test, the data are first clustered independently of the treatment group, using nonmetric clustering and the computer program RIFFLE (32). Because the RIFFLE analysis is naive

to treatment group, the clusters may, or may not correspond to treatment effects. To evaluate whether the clusters were related to treatment groups, whenever the clustering procedure produced four clusters for the sample points, the association between clusters and treatment groups was measured in a 4 x 4 contingency table, each point in treatment group i and cluster j being counted as a point in frequency cell ij. Significance of the association in the table was then measured with Pearson's X^2 test, defined as

$$\chi^2 = \sum_{ij} \frac{\left(N_{ij} - n_{ij}\right)^2}{n_{ij}}$$

where N_{ij} is the actual cell count and n_{ij} is the expected cell frequency, obtained from the row and column marginal totals N_{+j} and N_{i+} as

$$n_{ij} = \frac{N_{+j} N_{i+}}{N}$$

where $N = 24$ is the total cell count (33), and a standard procedure for computing the significance (probability) of χ^2 taken from (34).

REFERENCES

1. **National Research Council**. 1983. *Risk Assessment in the Federal Government.* National Academy Press, Washington, D.C.

2. **Suter, G.** 1993. A critique of ecosystem health: Concepts and indices. *Environ Tox. Chem.* in press.

3. **U. S. Environmental Protection Agency**. 1992. Framework for Ecological Risk Assessment. EPA/630/R-92/001.

4. **Suter II, G.W.** 1990. Endpoints for regional risk assessments. *Environmental Management.* 14:9-23.

5. **Hutchinson, G.E.** 1959. Concluding remarks. *Cold Spring Harbor Symposium on Quanitative Biology.* **22**:415-427.

6. **Hutchinson, G.E.** 1961. The paradox of the plankton. *Amer. Nat.* **95**:137-143.

7. **Richerson, P., R. Armstrong** and **C.R. Goldman**. 1970. Contemporaneous disequilibrium, a new hypothesis to explain the "paradox of the plankton". *Proc. Natl. Acad. Sci. U.S.A.* **67**:1710-1714.

8. **Tilman, D.** 1982. *Resource Competition and Community Structure.* Princeton University Press, Princeton, NJ, pp. 11-138.

9. **Landis, W.G.** 1986. Resource competition modeling of the impacts of xenobiotics on biological communities. In T.M. Poston and R. Purdy, eds., *Aquatic Toxicology and Environmental Fate: Ninth Volume, ASTM 921.* American Society for Testing and Materials, Philadelphia, PA, pp. 55-72.

10. **Conquest, L.L.** and **F.B. Taub.** 1989. Repeatability and reproducibility of the Standard Aquatic Microcosm: Statistical properties. In U.M. Cowgill and L.R. Williams, eds., *Aquatic Toxicology and Hazard Assessment: 12th Volume, ASTM STP 1027.* American Society for Testing and Materials, Philadelphia, PA, pp. 159-177.

11. **Kersting, K.** 1984. Development and use of an aquatic micro-ecosystem as a test system for toxic substances. Properties of an aquatic micro-ecosystem IV. *Int. Revue ges. Hydrobiol.* **69**:567-607.

12. **Kersting, K.** 1985. Properties of an aquatic micro-ecosystem V. Ten years of observations of the prototype. *Verh. Internat. Verein. Limnol.* **22**:3040-3045.

13. **Kersting, K.** 1988. Normalized ecosystem strain in micro-ecosystems using different sets of state variables. *Verh. Internat. Verein. Limnol.* **23**:1641-1646.

14. **Kersting, K.** and **R. van Wijngaarden.** 1992. Effects of Chloropyrifos on a microecosystem. **11**:365-372.

15. **Johnson, A.R.** 1988a. Evaluating ecosystem response to toxicant stress: a state space approach. *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP* 971, W.J. Adams, G.A. Chapman and W.G. Landis Eds., American Society for Testing and Materials, Philadelphia, pp. 275-285.

16. **Johnson, A.R.** 1988b. Diagnostic variables as predictors of ecological risk. *Environmental Management* **12**:515-523.

17. **Matthews, R.A., G.B. Matthews** and **W.J. Ehinger.** 1991. Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modeling.* **53**:167-187.

18. **Matthews, G.B., R.A. Matthews** and **B. Hachmoller.** 1991. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences.* **48**:2184-2190.

19. **Taub, F.B., A.C. Kindig** and **L.L. Conquest.** 1987. Interlaboratory testing of a standardized aquatic microcosm. In W.J. Adams, G.A. Chapman and W.G. Landis, eds., *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP 971.* American Society for Testing and Materials, Philadelphia, PA, pp. 385-405.

20. **Taub, F.B., A.C. Kindig, L.L. Conquest** and **J.P. Meador.** 1988. Results of the interlaboratory testing of the Standardized Aquatic Microcosm protocol. In G. Suter and M. Lewis, eds., *Aquatic Toxicology and Hazard Assessment: Eleventh Symposium, ASTM.* American Society for Testing and Materials, Philadelphia, PA.

21. **Taub, F.B.** 1988. Standardized aquatic microcosm - development and testing. *Aquatic Ecotoxicology* II.

22. **Taub, F.B.** 1989. Standardized aquatic microcosms. *Environm. Sci. Technol.* **23**:1064-1066.

23. **Kindig, A.C. , L.C. Loveday** and **F.B. Taub.** 1983. Differential sensitivity of new versus mature synthetic microcosms to streptomycin sulfate treatment. In W.E. Bishop, R.D. Cardwell and B.B. Heidolph, eds., *Aquatic Toxicology and Hazard Assessment: Sixth Symposium, ASTM 802.* American Society for Testing and Materials, Philadelphia, PA, pp. 192-203.

24. **Landis, W.G., M.V. Haley** and **N.A. Chester.** 1993. The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In W.G. Landis, J. Hughes and M. Lewis., eds., *In press, Environmental Toxicology and Risk Assessment: First Volume, ASTM STP -1179.* American Society for Testing and Materials, Philadelphia, PA.

25. **Matthews, G.** and **R. Matthews.** 1991. A model for describing community changes. *Proceedings,* Conference on Pesticides in Natural Systems, EPA 910/9-91-001, United States Environmental Protection Agency Region 10, Corvallis, OR.

26. **Landis, W.G., R.A. Matthews, A.J. Markiewicz, N.J. Hough** and **G.B. Matthews.** 1993. Multivariate analysis of the impacts of turbine fuel using a standard aquatic microcosm toxicity test. *J. Env. Sci.* In Press.

27. **Giddings, J.M., B.R. Parkhurst, C.W. Hehrs** and **R.E. Millemann.** 1980. Toxicity of a coal liquefaction product to aquatic organisms. *Bull. Environ. Contam. Toxicol.* **25**:1-6.

28. **Good, I.J.** 1982. An index of separateness of clusters and a permutation test for its significance. *J. Statist. Comp. Simul.* **15**:81-84.

29. **Smith, E.P., K.W. Pontasch** and **J. Cairns, Jr.** 1990. Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.* **24**:507-514.

30. **Matthews, G.B.** and **R.A. Matthews.** 1990. A model for describing community change. In *Pesticides in Natural Systems: How Can Their Effects Be Monitored? Proceeding of the Conference,* Environmental Research Laboratory/ORD, Corvallis, OR, EPA 9109/9-91/011.

31. **Noreen, E.W.** 1989. *Computer Intensive Methods for Testing Hypotheses.* Wiley-Interscience, New York, NY.

32. **Matthews, G.B.** and **J. Hearne.** 1991. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **13**:175-184.

33. **Flenberg, S.E.** 1985. *The Analysis of Cross-Classified Categorical Data.* MIT Press, Cambridge, MA.

34. **Press, W.H., B.P. Flannery, A.A. Teukolsky** and **W.T. Vetterline.** 1990. *Numerical Recipes in C, the Art of Scientific Computing.* Cambridge University Press, New York, NY.

Tables

Table 1. Important Variables Ranked By Nonmetric Clustering For Each Sampling Date For The Jet-A SAM Toxicity Test. Some variables such as *Ankistrodesmus* were consistently important in determining group clusters throughout the experiment. Some of the variables such as **Ostracod** and Philodina were more important in the latter stages of the experiment. The order of importance of the variables often changed from day to day, with no one variable being common to each sampling date. The variables used as part of the overall analysis were: Anabaena, Ankistrodesmus, Chlamydomonas, Chlorella, Daphnia (Ephipia, Small Daphnia, Medium Daphnia, Large Daphnia), Hypotricha, Lyngbya, Miscellaneous sp., Ostracod (Cyprinotus), Philodina (Rotifer), Scenedesmus, Selenastrum, Stigeoclonium, and Ulothrix.

Day	Important Variables in Determining Clusters in Rank Order
11	M. Daphnia, Chlorella, Chlamydamonas, Ulothrix, S. Daphnia,Selanastrum,Scenedesmus
14	S. Daphnia, M. Daphnia-Selenastrum[1], Chlamydamonas, Chlorella, L. Daphnia, *Ankistrodesmus*
18	*Ankistrodesmus*, S. Daphnia, Chlorella, Chlamydamonas, Selanstrum, L. Daphnia
21	*Ankistrodesmus*, S. Daphnia, L. Daphnia-M. Daphnia, Scenedesmus
25	Scenedesmus, S. Daphnia, L. Daphnia, Chlorella, Philodina-M. Daphnia
28	*Ankistrodesmus*, L. Daphnia, Scenedesmus
32	S. Daphnia, M. Daphnia, *Ankistrodesmus*, Chlorella
35	*Ankistrodesmus*
39	M. Daphnia-Selenastrum, **Ostracod**-*Ankistrodesmus*
42	M. Daphnia, **Ostracod**, Scenedesmus
46	Scenedesmus, *Ankistrodesmus*, S. Daphnia. M. Daphnia
49	Chlorella, Philodina, *Ankistrodesmus*, Lyngbya
53	*Ankistrodesmus*, **Ostracod**, Chlorella
56	M. Daphnia-Scenedesmus, *Ankistrodesmus*, Lyngbya
60	Lyngbya, M. Daphnia, Philodina, Chlorella
63	Chlorella, *Ankistrodesmus*, Philodina, **Ostracod**

[1] Hyphen between variables denotes equal rank

Table 2. Variable According to Success in Determining Clusters as Defined by Nonmetric Clustering in the Jet-A SAM Experiments. Variables such as Ankistrodesmus and the Daphnia classes were important in the course of this study. Reliance on even these two variables would have been misleading in the determination of the second oscillation.

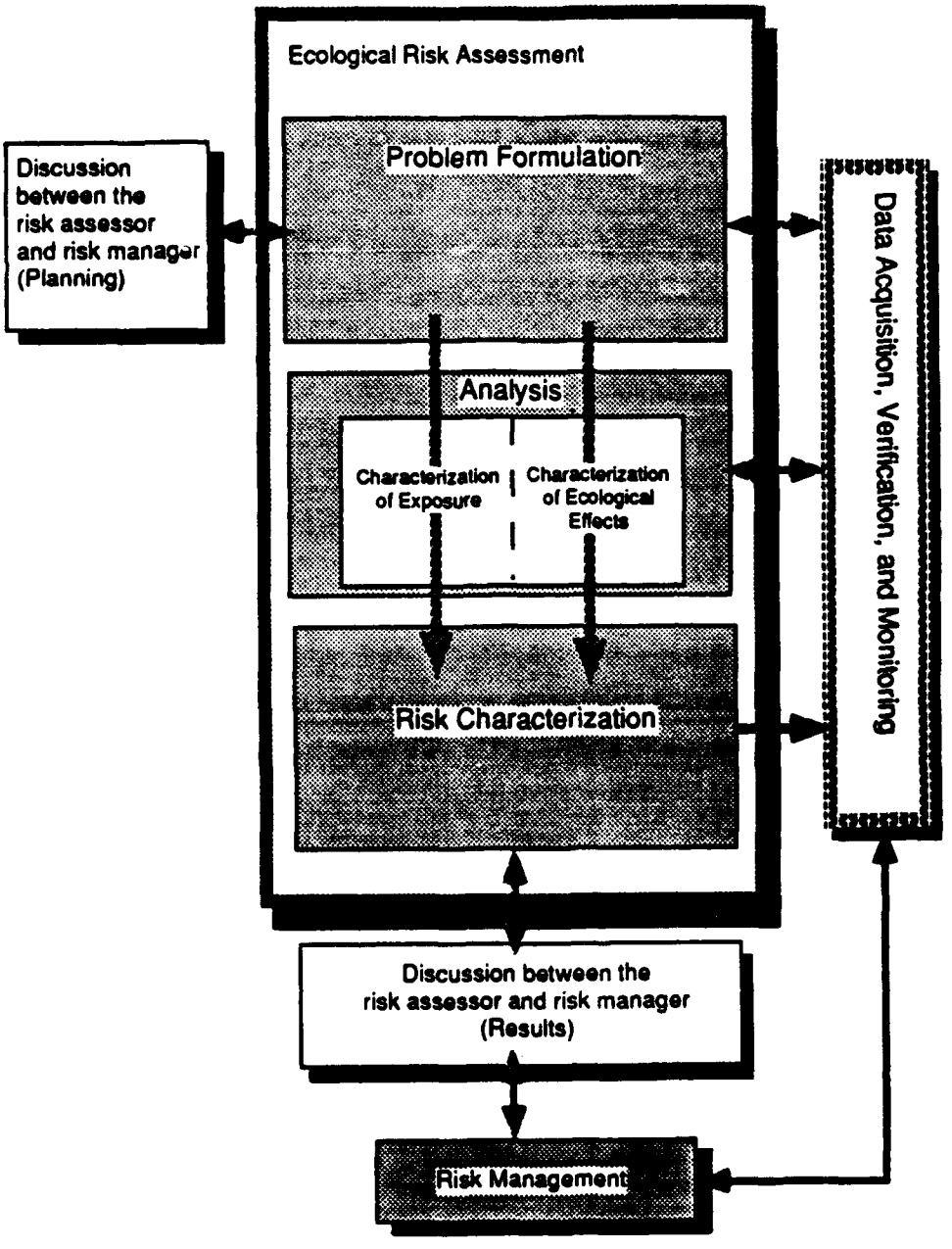
Variable	Ranked
Ankistrodesmus	12
M. Daphnia	11
Chlorella	9
Scenedesmus	7
S. Daphnia	6
L. Daphnia	5
Ostracod	4
Philodina	4
Selenastrum	4
Lyngbya	3
Ulothrix	1

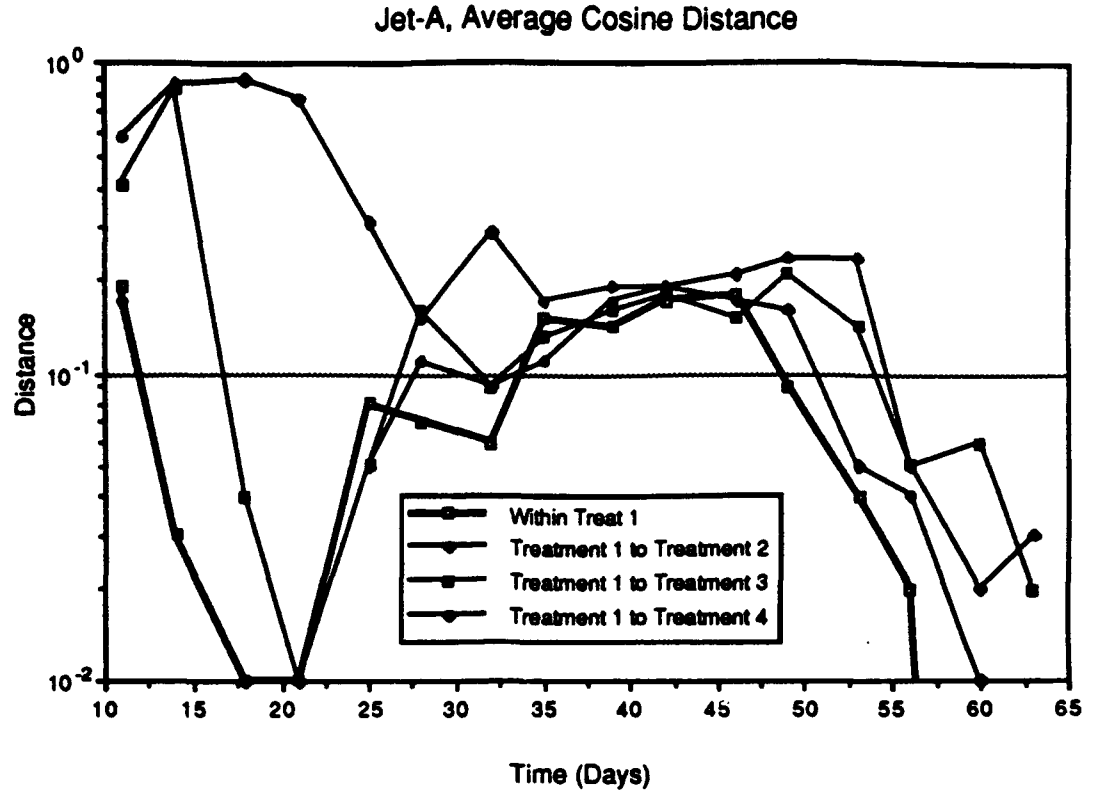
Figures

Figure 1. Schematic of the Framework for Ecological Risk Assessment (3). Especially important in the interaction between exposure and hazard and the inclusion of a data acquisition, verification and monitoring component. Multivariate analyses will have a major impact upon the selection or assessment and measurement endpoints as well as playing a major role in the data acquisition, verification and monitoring phase.

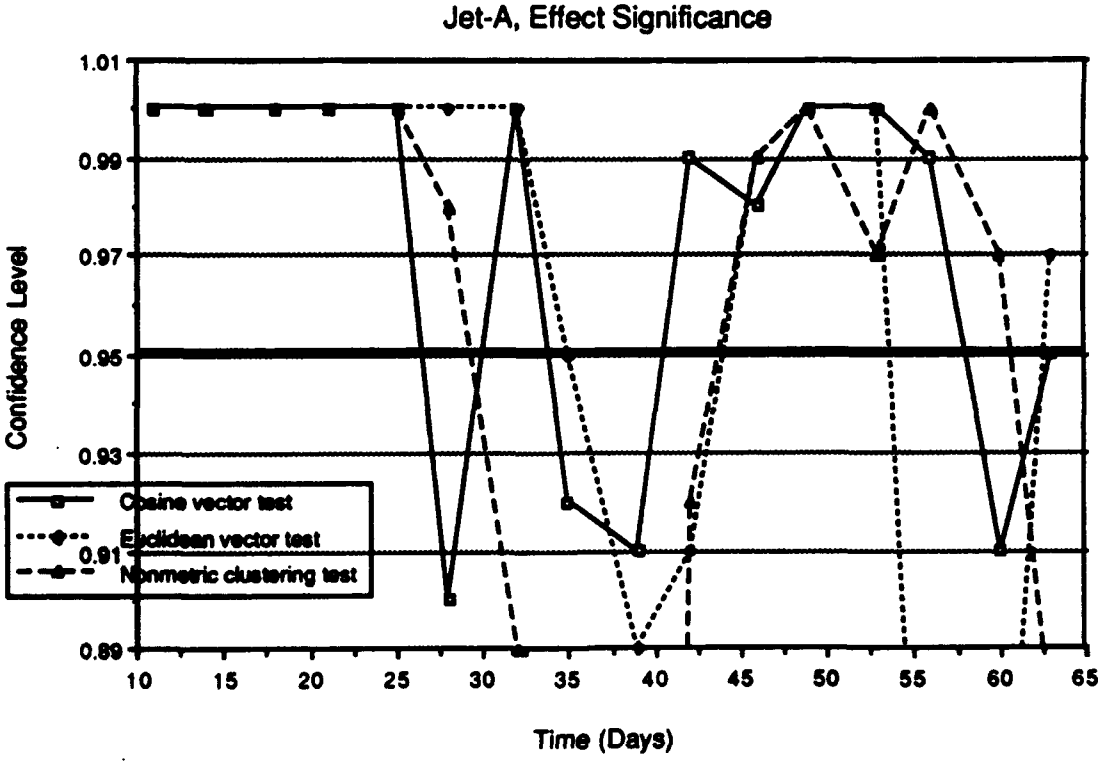
Figure 2. Multivariate analysis of the impact of Jet-A in the SAM test system. Figure 2A shows the Cosine distance from the control group to each of the treatments for each sampling day. Note that large differences are apparent early in the SAM. During the middle part of the 63 day experiment the distances between the replicates of Treatment 1, the control group, is as large as the distances to the treatment groups. However, later in the experiment the distances from the dosed microcosms to the control again increase. Significance levels of the three multivariate statistical tests for each sampling day are presented in Figure 2B. Note that there are two periods, early and late ones, where the clustering into treatment groups is significant at the 95 percent confidence level or above.

Figure 3. *Measures of distance between clusters.* Two of the commonly used measures of separation of clusters in a n-dimensional space are the cosine of the angle and the vector distance. Each method has advantages and disadvantages. In order to visualize the data as accurately as possible several measures should be employed.

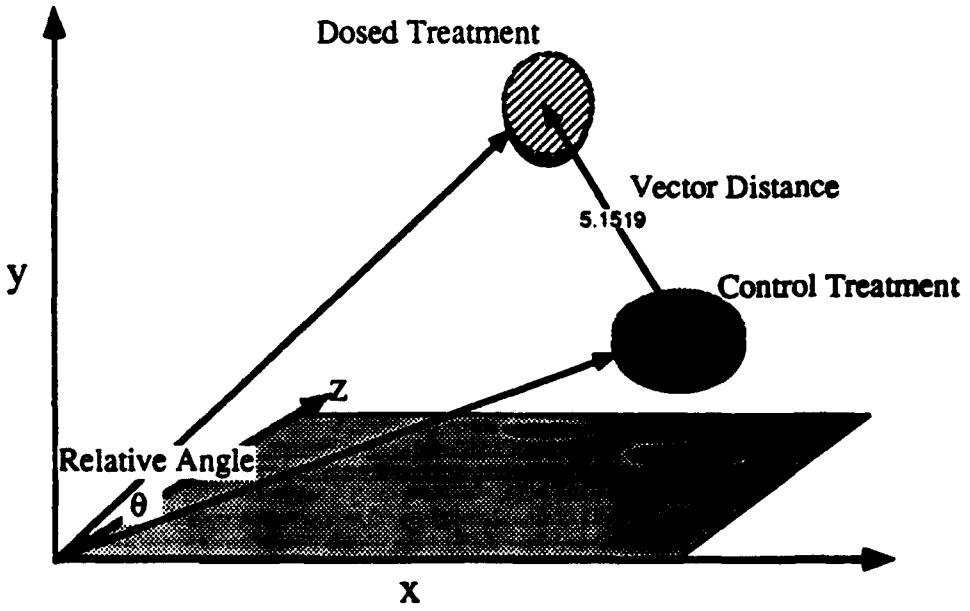

Ecological Risk Assessment
Discussion between the risk assessor and risk manager (Planning)
Problem Formulation
Analysis
Characterization of Exposure
Characterization of Ecological Effects
Risk Characterization
Data Acquisition, Verification, and Monitoring
Discussion between the risk assessor and risk manager (Results)
Risk Management


Jet-A, Average Cosine Distance
Distance
10^0
10^{-1}
10^{-2}
10
15
20
25
30
35
40
45
50
55
60
65
Time (Days)
Within Treat 1
Treatment 1 to Treatment 2
Treatment 1 to Treatment 3
Treatment 1 to Treatment 4

A


Jet-A, Effect Significance
Confidence Level
1.01
0.99
0.97
0.95
0.93
0.91
0.89
10
15
20
25
30
35
40
45
50
55
60
65
Time (Days)
Cosine vector test
Euclidean vector test
Nonmetric clustering test

B


Dosed Treatment
Vector Distance
5.1519
Control Treatment
y
z
Relative Angle
θ
x

Comparison of the Community Dynamics of the Standardized Aquatic Microcosm and Mixed Flask Culture Microcosm Toxicity Tests with Suggestions for Design Criteria in Multispecies Toxicity Testing

Wayne G. Landis*, April J. Markiewicz and Robin A. Matthews

Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies

Western Washington University, Bellingham, WA 98225.

Michael J. Roze and Geoffrey B. Matthews

Computer Science Department

Western Washington University, Bellingham, WA 98225.

Running Title: Multispecies Toxicity Testing Comparison

*To whom correspondence should be addressed

Corresponding Author: Wayne G. Landis, Institute of Environmental Toxicology and Chemistry, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA 98225.

Phone: 206-650-6136, Fax: 206-650-7284, Internet: landis@henson.cc.wwu.edu

Abstract -- The water soluble fraction of the turbine fuels Jet-A and JP-4 have been examined as stressors for two microcosm protocols, the standardized aquatic microcosm (SAM) and the mixed flask culture (MFC). The SAM is a 3 L system inoculated with standard cultures of algae, zooplankton, bacteria, and protozoa. In contrast, the MFC is 1 L and is inoculated with a complex mixture of organisms derived from a natural source. Analysis of the organism counts and physical data were conducted using conventional and newly derived multivariate nonmetric clustering methods. In both the SAM and MFC test systems, species numbers and other variables that determined clusters varied among sampling dates. Compared to the larger yet simpler SAM system, the MFC exhibits more violent and erratic dynamics. The variability in the responses may be due to at least three factors, the relatively small size yet high species number of the MFC relative to the SAM, the inadequacy of the cross inoculation procedure to set initial conditions, and the almost two fold increase in surface area/volume of the MFC system. Although both experiments are performed as specified, recovery is not apparent using nonmetric clustering and association analysis as well as other more conventional means. Suggestions to improve the resolution of multispecies toxicity tests include: sampling variables that represent the metabolism and structure of the procaryotic community, eliminate cross inoculation and accept the heterogeneity of the replicates, and use methods that explore the dataset in a multidimensional space, and finally accept the nonequilibrium nature of multispecies toxicity tests as representative of natural systems.

Key Words: Multispecies toxicity test, Standardized Aquatic Microcosm, Mixed Flask Culture, Non-metric clustering and association analysis, non-equilibrium dynamics

INTRODUCTION

There has been a renewed interest regarding the use of multispecies toxicity tests and in the evaluation of changes to ecological communities. The recent decision to limit the use of field and multispecies tests in pesticide registration [1] has sparked a great deal of debate about the appropriateness of this level of toxicity evaluation. Although many factors contributed to the action, apparently the field and pond mesocosm tests that were conducted as part of the registration process did not contribute to the evaluation of risk of pesticides in a timely and cost effective manner. This is in spite of a number of available methods and analysis techniques.

Over the last 15 years a variety of multispecies toxicity tests have been developed. Multispecies toxicity tests are usually referred to as microcosms or mesocosms, although a clear definition of the size or complexity to distinguish these terms has not been put forth. Multispecies toxicity tests range from approximately 1 L (e.g., mixed flask cultures) to thousands of liters, as in the case of the pond mesocosms used in pesticide registration testing. A recent review by Gearing [2] listed eleven freshwater artificial stream methods, 22 laboratory freshwater microcosms ranging from .1 to 8,400 liters, 18 outdoor freshwater microcosms ranging from 8 to 18,000,000 liters, and even larger numbers of marine systems.

The Mixed Flask Culture (MFC) and the Standardized Aquatic Microcosm (SAM) were initially developed to examine the population dynamics, food-trophic level interactions, and the relationships between community structure and community function that are not possible with single species toxicity tests [3]. Complexity in the sense of total species numbers and total possible interactions are traded for the purpose of establishing generic structural and functional processes. An underlying assumption or even hope of this approach was that all ecosystems display the similar patterns and behaviors in their structural and functional relationships. Perhaps there exists universal ecosystem properties and universal patterns of responses to stress [3].

Previous comparisons of the MFC and SAM by Stay et al. [4] demonstrated that coefficients of variation within these systems were low with chemical measurements and much higher with structural parameters such as organism counts. Large changes due to the toxicant stress did increase the coefficient of variation. In order to test the accuracy of these systems comparisons to field tests of several pesticides were made. Stay et al. did conclude that these systems did accurately reflect field tests conducted with atrazine, fluorene and chlorpyrifos. Since the publication of this report, new methods of evaluating the dynamics of multispecies systems have been developed.

One of the major difficulties in the evaluation of multispecies toxicity tests has been the difficulty in the analysis of the large data set on a level consistent with the goals of the toxicity test. Typically, the goals of the toxicity test are:

- to detect changes in the population dynamics of the individual taxa that would not be apparent in single species tests;

- to examine the fate of the introduced toxicant; and

- to detect community-level differences that are correlated with treatment groups thereby representing a deviation from the control group.

A number of methods have been developed to attempt to satisfy the goals of multispecies toxicity testing. Analysis of variance (ANOVA) is the classical method to examine single variable differences from the control group. However, because multispecies toxicity tests generally run for weeks or even months, there are problems with using conventional ANOVA. These include the increasing likelihood of introducing a Type II error (accepting a false null-hypothesis), temporal dependence of the variables, and the difficulty of graphically representing the data set. Conquest and Taub [5] developed a method to overcome some of the problems by using intervals of non-significant difference (IND). This method

corrects for the likelihood of Type II errors and produces intervals that are easily graphed to ease examination. The method is routinely used to examine data from SAM toxicity tests, and it is applicable to other multivariate toxicity tests. The major drawback is the examination of a single variable at a time over the course of the experiment. While this addresses the first and perhaps second goal in multispecies toxicity testing, listed above, it ignores the third. In many instances, community-level responses are not as straightforward as the classical predator/prey or nutrient limitation dynamics. The interactions among the various measured parameters had to be gleaned by the investigator examining each variable independently.

Multivariate methods have proved promising as a method of incorporating all of the dimensions of an ecosystem. One of the first methods used in toxicity testing was the calculation of ecosystem strain developed by Kersting [6,7,8] for a relatively simple (three species) microcosm. This method has the advantage of using all of the measured parameters of an ecosystem to look for treatment-related differences. At about the same time, Johnson [9,10] developed a multivariate algorithm using the n-dimensional coordinates of a multivariate data set and the distances between these coordinates as a measure of divergence between treatment groups. Both of these methods have the advantage of examining the ecosystem as a whole rather than by single variables, and can track such processes as succession, recovery and the deviation of a system due to an anthropogenic input.

However, a major disadvantage of both these methods, and of many conventional multivariate methods, is that all of the data are often incorporated without regard to the units of measurement or the appropriateness of including all variables in the analysis. It can be difficult to combine variables such as pH, with units ranging from 0-14, with the numbers of bacterial cells per ml, where low numbers are in the 10^6 range, to say nothing of the conceptual difficulties of adding pH units to counts. Similarly, random variables (i.e., variables with no treatment-related response) indiscriminately incorporated into the analysis may contribute so much noise that they overshadow variables that do show treatment-related effects. We

have implemented new techniques to the analysis of patterns in ecological datasets, nonmetric clustering and association analysis.

Unlike the more conventional multivariate statistics, nonmetric clustering is an outgrowth of Artificial Intelligence (AI) and a tradition of conceptual clustering. In this approach, an accurate description of the data is only part of the goal of the statistical analysis technique. Equally important is the intuitive clarity of the resulting statistics. For example, a linear discriminant function to distinguish between groups might be a complex function of dozens of variables, combined with delicately balanced factors. While the accuracy of the discriminant may be quite good, use of the discriminant for evaluation purposes is limited because humans cannot perceive hyperplanes in highly dimensional space. By contrast, conceptual clustering attempts to distinguish groups using as few variables as possible, and by making simple use of each one. Rather than combining variables in a linear function, for example, conjunctions of elementary "yes-no" questions could be combined: species A greater than 5, species B less than 2, and species C between 10 and 20. Numerous examples throughout the artificial intelligence literature have proven that this type of *conceptual* statistical analysis of the data provides much more useful insight into the patterns in the data, and is often more accurate and robust. Delicate linear discriminants, and other traditional techniques, chronically suffer from overfitting, particularly in highly dimensioned spaces. Conceptual statistical analysis attempts to fit the data, but not at the expense of a simple, intuitive result. Patterns detected by the clustering are then tested against the hypothesized pattern using association analysis. A more detailed description of nonmetric clustering and association analysis has been published [11].

An additional advantage of using clustering methodologies in the comparison of two experimental methods is the ability to compare the results at a fundamental level. The question can be simply put; can differences in the treatment groups be detected and for what period of time? Since the nonmetric clustering procedure ranks variables in terms of importance, these rankings can be examined for patterns indicating similarities or differences in metabolic processes or structural composition.

In this report we would like to focus on a comparison between the Standardized Aquatic Microcosm (SAM) as developed by Taub [12] and ASTM E 1366-91 [13], and the Mixed Flask Culture microcosm method as developed by Leffler [14] and later modified by Shannon et al. [15]. Over the last three years our research group has examined the toxicity of turbine fuels using the Standardized Aquatic Microcosm and the Mixed Flask Culture. In the process of using these experimental methods we have also used a variety of conventional and novel analysis techniques to examine the responses of these systems to the turbine fuels. This report is a comparison of the methods in light of our recent research findings [16,17] and our clustering and artificial intelligence methodologies. We also suggest possible modifications to these methods, multispecies toxicity tests in general, and the analysis of community level effects.

EXPERIMENTAL METHODS

Chemicals

All chemicals used in the culture of the organisms for the Standardized Aquatic Microcosm and in the preparation of the microcosm medium, T82MV, were *reagent grade or* as specified by the ASTM and USEPA protocols. Individual hydrocarbon reference standards, that were used to identify and quantify the water soluble components in the jet fuels, were purchased from the Alltech Chemical Company (Deerfield, IL), were certified to 99+% purity and A.C.S. spectrophotometric grade. The ASTM D3710 Qualitative Calibration Mix and the Qualitative Reference Reformate Standard were purchased from Supelco Chromatography Products (Bellefonte, PA). All standards were prepared in pesticide residue grade, A.C.S. specification hexane or carbon disulfide, purchased from VWR Scientific (Seattle, WA).

The jet fuel formulations, used in the two microcosms, were Jet-A, used in the commercial aircraft, and JP-4, used in the U.S. Air Force military aircraft. Jet-A is refined by Chevron and was provided locally by Fliteline Services of Bellingham, Washington. JP-4 was supplied by the U.S. Air Force Toxicology Laboratory at Wright Patterson Air Force Base in Ohio. The samples were collected in two liter fuel cans

from in-line quality assurance/quality control valves, sealed on site, lot shipment recorded and transported to the laboratory, using in-place chain-of-custody procedures.

Water Soluble Fractions

The water soluble fraction (WSF) of Jet-A and JP-4 were prepared in glassware washed in nonphosphate soap, rinsed, soaked in 2N HCl for at least one hour, rinsed ten times with distilled water, dried and finally autoclaved for 30 minutes. Microcosm medium, T82MV, was substituted as the diluent for the water fraction of the WSF. One liter separatory funnels were used as mixing chambers to prepare the 100% WSF due to the control of venting built up gases during the mixing process; the minimal head space that would prevent potential loss of volatiles; and ease of separating and removing the hydrocarbon saturated water fraction from the liquid fuel fraction.

Twenty-five mL of the appropriate jet fuel were added to each one liter separatory funnel containing one liter of sterile, fresh T82MV medium and mixed by agitating the separatory funnel con 's vigorously for five minutes, slowly releasing built up pressure when necessary; then allowing the contents to stand undisturbed for fifteen minutes, and repeating this procedure until a total time of one hour had elapsed. The separatory funnel and its contents were then allowed to remain undisturbed for twelve hours at 20°C, to maximize the saturation of the T82MV with the water soluble components in the jet fuel.

After twelve hours, the T82MV/100% water soluble fraction of jet fuel mixture was slowly drained from the separatory funnel, being careful to leave behind the final 100 mL in direct contact with the jet fuel layer; to avoid incorporating any jet fuel emulsion into the water soluble fraction.. The 100% WSF was placed directly into clean, sterile one liter amber glass bottles and capped with Teflon-lined screw caps. The 100% WSF was used within twelve hours of preparation.

Gas Chromatography of WSF

A Tekmar LSC 2000 Purge and Trap (P&T) concentrator system in tandem with a Hewlett Packard 5890A Gas Chromatograph with a Flame Ionization Detector (FID) was used for the analysis of all microcosm samples and standards. Instrument blanks and deionized distilled water blanks are used to verify the P&T and GC columns cleanliness prior to analysis of samples. A 5 mL gas tight Teflon Luer lock syringe was used to remove a 3.5 mL sample and inject it into the 5 mL sparger where the sample was purged with pre-purified nitrogen gas for eleven minutes and dry-purged for four minutes. Volatile hydrocarbons, purged from the sample and collected on the Tenax/Silica Gel column, were desorbed at 180°C directly onto the gas chromatograph SPB-5, 30 m x 0.53 mm ID 1.5 µm film, fused silica capillary column. The GC column was programmed to hold at 35°C for two minutes, increase to 225°C at 12°C/min and hold at that temperature for five minutes. A Spectra-Physics 4290 Integrator recorded the FID signal output of the volatile hydrocarbons, separated and eluted from the column by molecular weight and boiling point. A comparison was then made of the sample chromatograph peak retention times and area under the peak curve to n-paraffin and aromatic chromatograph reference standards, prepared and analyzed under the same conditions, for sample concentration determinations.

Standardized Aquatic Microcosm (SAM) Protocol

The 63-day SAM protocol previously has been described [13]. Briefly, the microcosms were prepared by the introduction of ten algal, four invertebrate, and one bacterial species into 3 L of sterile, chemically defined medium. Test containers were 4 L glass jars, containing an artificial sediment of 200 g silica sand, 0.5 g of cellulose, and 0.5 g of ground chitin and filled with 3 L of the T82MV medium. The jars are autoclaved and immersed in a water bath to a point above the level of sand during autoclaving. This procedure helps prevent breakage of the jars and subsequent loss of replicates. The numbers of organisms, dissolved oxygen (DO) and pH were determined twice weekly. The laboratory environmental conditions were maintained at a temperature of 20°C ± 2°; illumination was 79.2 μEm^{-2} sec^{-1} photosynthetically active radiation, with a range of 78.6 - 80.4 μEm^{-2} sec^{-1}; and a 12:12 day/night cycle.

Two major modifications were made to the SAM protocol. The first was the means of toxicant delivery. On day 7, 450 mL were removed from each container using an autoclaved, 100 mL capacity basting tube, with a sterile square of 100 mesh Nitex® tied over the opening to prevent the removal of the organisms. The 100% WSF stock material was then combined with fresh, sterile T82MV and added in appropriate amounts to produce concentrations of 0, 1, 5 and 15 percent WSF for the four treatment groups. After toxicant addition the final volume was adjusted to 3 L. All graphs and statistical analysis start with the first sampling day, day 11. The second modification was the substitution of *Tetrahymena thermophila* BIV for the hypotrichous ciliate used in past experiments. The results presented below demonstrate the suitability of the Tetrahymena for inclusion in the protocol. The microcosms were monitored for structural parameters, with subsamples removed from each microcosm and counts of population densities made for all species, on Tuesdays and Fridays, for the duration of the 63 day experiment.

Mixed Flask Culture Microcosm Protocol

Construction and implementation of the 60-day Mixed Flask Culture microcosm experiment was conducted to the specifications described in the USEPA document PB89-221295. In brief, natural occurring assemblages of aquatic organisms were collected from local streams and lakes, brought back to the laboratory, placed in a 50 L aquarium containing the same, chemically defined sterile medium T82MV used in the ASTM Standardized Aquatic Microcosm [13] and allowed to reassemble and restructure during a three month equilibration period. Laboratory environmental conditions were maintained at 20° ± 2°C, light intensity at 80 ± 2 $\mu Em^{-2}s^{-1}$, and a photoperiod of 12 hours light and 12 hours dark. At the end of three months, the resulting co-adapted community was subsampled, with 50 mL removed and inoculated into each of the thirty, cleaned and acid washed 1 L beakers containing 50 g acid-washed white silica sand, 15 µg $NaHCO_3$ as an additional carbon source, and 900 mL of freshly made, sterile T82MV medium. The beaker microcosms were then placed, in a Puffer-Hubbard CEC 50LTP Environmental Chamber, with the environmental conditions set to an isothermal day/night temperature of 20° ± 2°C, illumination at 80 ± 2 $\mu Em^{-2}s^{-1}$, and a photoperiod of 12 hours light and 12 hours dark.

The microcosms were allowed to equilibrate for six weeks during which time they were cross-inoculated once a week to minimize divergence; re-inoculated once a week to ensure a more uniform distribution of organisms among the beakers; and rotated within the environmental chamber twice a week to minimize potential light and temperature variations. Cross-inoculation and re-inoculation procedures were combined to minimize the disturbance to the microcosms. After the six week equilibration period, the microcosms were examined individually to verify that each contained the specified minimum functional groups; two species of unicellular green algae; one species of nitrogen fixing blue-green algae; one species of filamentous green algae; one species of herbivorous grazer; one species of benthic detritivore; bacteria; and protozoans. A total of twenty-four microcosms were selected, based on minimum variance from the mean for pH and DO, then randomly numbered and assigned to four treatment groups, each containing six replicates, with one treatment group of six replicates (Treatment 1) serving as the reference or non-dosed microcosms. Test material was added on day 0 by stirring each microcosm, removing 150 mL from each container, using an autoclaved, 100 mL capacity basting tube, with a sterile square of 100 mesh Nitex® tied over the opening to prevent the removal of the organisms, and then adding appropriate amounts of the 100% WSF stock material to produce concentrations of 0, 1, 5 and 15 percent WSF. After toxicant addition the final volume was adjusted to 1 L. All graphs and statistical analysis start with the first sampling day, day 4. The microcosms were monitored for structural parameters, with subsamples removed from each microcosm and counts of population densities made for all species, on Tuesdays and Fridays, for the duration of the experiment. The duration of the experiments was 56 days for the JP-4 and 77 days for Jet-A.

Data Analysis and Visualization

Nonmetric clustering and association analysis. All data were recorded onto standard computer entry forms and checked for accuracy. The data was then keyed into standard data recording spreadsheets and checked for accuracy. Parameters calculated included the concentrations of each of the species, DO, DO gain and loss, net photosynthesis/respiration ratio (P/R), pH, algal species diversity, algal biovolume, and biovolume of available algae. Note that algal biovolume, algal species diversity and available algae are

all derived variables based on the algal counts. The net photosynthesis/respiration ratio is not derived using ^{14}C methods but by comparing oxygen concentrations before lights on, at the end of the photosynthetic period, and then at the next morning, as specified in the standard protocol. Photosynthesis/respiration ratio was the variable used during the analysis to incorporate these measurements.

The multivariate clustering methods used in the comparison was nonmetric clustering and association analysis. This method and its application to ecological datasets has been previously described [16,17,18]. In the nonmetric clustering and association test, the data are first clustered independently of the treatment group, using nonmetric clustering and the computer program RIFFLE [11]. Because the RIFFLE analysis is naive to treatment group, the clusters may, or may not correspond to treatment effects. To evaluate whether the clusters were related to treatment groups, whenever the clustering procedure produced four clusters for the sample points, the association between clusters and treatment groups was measured in a 4 x 4 contingency table, each point in treatment group i and cluster j being counted as a point in frequency cell ij. Significance of the association in the table was then measured with Pearson's X^2 test, defined as

$$\chi^2 = \sum_{ij} \frac{(N_{ij} - n_{ij})^2}{n_{ij}}$$

where N_{ij} is the actual cell count and n_{ij} is the expected cell frequency, obtained from the row and column marginal totals N_{+j} and N_{i+} as

$$n_{ij} = \frac{N_{+j} N_{i+}}{N}$$

where $N = 24$ is the total cell count, and a standard procedure for computing the significance (probability) of χ^2 taken from Press e al. [19].

Projections of ecosystem dynamics, space-time worms. One way of visualizing this day-to-day change in a two-dimensional projection of the data is with a three-dimensional, interactive computer graphic of the resulting space-time "worm": the cylindrical surface generated by the two data dimensions and time. Three-dimensional space-time worms can depict two-dimensional dynamics of ecological systems and allow better comparisons than traditional, one-dimensional graphs. The generation of these projections is typically done using two of the parameters selected by the NMCAA as important in the determination of clusters. Two axes are used to generate the coordinates of the treatment group and a circle proportional to the dispersion among the six replicates is drawn. The third axis, time, is then added and the resulting three dimensional structure is projected using Renderman for the NeXT STEP operating system. The resultant projection is then rotated in real time so that a perspective providing the best viewing of the dynamics of the system can be selected.

RESULTS

The results presented here are limited those associated with the nonmetric clustering results for each of the experiments. More extensive results have been published for the Jet-A and JP-4 SAM experiments [16,17, respectively].

In all of the experiments described in this report, the material derived from the jet fuel WSF has degraded by day 30 of the experiment. Although numerous peaks are present, by the mid point of the experiment little is apparently left in the water column. The degradation of these materials is extensively analyzed by Markiewicz [20].

The significance level of the nonmetric clustering is discerning treatment groups in the Jet-A and JP-4 SAM experiments is depicted in Figure 1. In both experiments the treatment groups were distinguishable from dosing until approximately day 25-30. A period where the treatments were not identifiable at the .95 or even .90 significance levels was followed by another clustering according to dose. The trend is more apparent in the Jet-A data as opposed to the JP-4 data.

The nonmetric clustering was done in two ways for the Jet-A and JP-4 MFC experiments. The first set of data included identification of each of the algae and other organisms as much as possible to individual genera or even species. However, since some groups, such as the ciliates, are difficult to identify taxonomically there are lumped into categories such as total ciliates. A second nonmetric clustering and association analysis were performed on a data set that lumped individual based data into categories such as Total Algae. The results of these analyses are presented in Figure 2. In both jet fuel experiments, there did not seem to be an increase in sensitivity. In some cases days that were significant did not correspond. In the comparisons that follow, both analysis sets are used.

The results of the Jet-A and JP-4 MFC tests are portrayed in Figure 3. Whether individual or total counts are used, the Jet-A results exhibit a greater number of points above both the .90 and .95 levels of significance. In the Jet-A experiment some of these points are found primarily in the later stages of the experiment, some of these after the normal 56 day run allocated for the JP-4 MFC experiment. This pattern is most clearly shown in the analysis using the individual counts. As in the SAM experiments, JP-4 apparently has less impact on the system than Jet-A.

A comparison of the SAM and MFC experiments to demonstrate effects as in the discrimination of treatment groups as detected by NMCAA is presented in Figure 4. The individual count analysis of the MFC is used since this is similar to the approach used in the SAM experiments. Since dosing with a toxicant occurs on different experimental days in the two protocols, the x axis has been adjusted to days since toxicant introduction. In both the comparisons, the MFC demonstrates a more erratic pattern than

does the SAM experiment. However, in the Jet-A experiments, an early and late resolution into treatment groups seems to be a common pattern. Such a pattern is not readily discernible in the JP-4 experiments.

Another advantage of the NMCAA is that it ranks variables in order of contribution to the establishment of clusters. Table 1 lists the important variables for each of the sampling days for both SAM experiments. Variables Ostracod has been put in boldface and Philodina underlined so that they can be readily identified in the lists. In the Jet-A and JP-4 SAM experiments these variables do not become important in discriminating clusters until the latter half of the experiment. As shown in Table 2, this is in contrast to the variables listed for the MFC experiments with the same toxicants. Ostracods are again boldface, two species were present, and Philodina was replaced by Rotifers and underlined to reflect the diversity of this group. In the case of the MFC experiments, these variables do not seem to reveal any particular pattern in the list of important variables.

The space-time worm projections for the Jet-A SAM and MFC experiments are depicted in Figure 5. We have used the Jet-A SAM and MFC experiments as examples. Two attribute axes are used in these projections, Ostracods and Ankistrodesmus. Ostracods and Ankistrodesmus are selected by the NMCAA as an important variable in both the SAM and MFC experiments. Ostracods are important in describing the treatment clusters in the second half of the SAM experiment, Ankistrodesmus is often an important variable in the beginning. Both Ankistrodesmus and Ostracods are often selected as important variables throughout the Jet-A MFC experiment. Although the systems are established in quite different manners, the patterns qualitatively are similar. Each has an initial separation followed by a convergence of the systems. The convergence is then followed by another differentiation. In the Jet-A SAM the initial divergence is quite large, reflecting the algal bloom due to the sensitivity of the grazers to the Jet-A WSF. In both experiments the 4 groups are still apparent by the end of the experiment. Even after the degradation of the toxicant, the systems have as an identity the toxicant treatments. The projections

provide additional confirmatory evidence that similar patterns occur even among dissimilar experimental programs.

DISCUSSION

Comparison of the SAM and MFC Protocols

The experimental designs (Table 3) of the two methods reveal a great deal of similarity. The numbers of groups and the replicates in each group are identical with a total of 24 experimental units available for analysis. The reinoculation of the SAM with algae and other taxa to simulate migration during the course of the experiment is not performed in the MFC. The greatest difference in the designs is the fact that the SAM system is inoculated with set amounts of organisms, minimizing historical inputs before the introduction of the toxicant. In the MFC protocol, a naturally derived inoculum is used. This inoculum is typically a combination of several collections and a three month maturation period occurs before samples are withdrawn for the test procedure. As the experimental units are constructed, a maturation period of 6 weeks is allowed with cross inoculation among the experimental units performed. Cross inoculation stops at the time of toxicant addition. This method allows for a greater number of species, many rare, and also sets each unit with it own historical identity.

In the physical construction of the microcosm units (Table 4) the systems are again similar. Total volume of the SAM is maintained at 3 K while the MFC is 950 mL of media. Not only is there less volume in the MFC, but a calculation of the surface of the container to volume ratio indicates that the MFC has 1.5 times the surface to volume ratio of the SAM method. Organisms and fate processes that are located on the glass surface and sediment are likely to occur at different rates in two systems.

The types of measurements taken as part of the SAM and MFC protocols are similar (Table 5). The biggest difficulty and difference is that in the MFC, with its larger number of species, it is difficult to identify the organisms to species level within a reasonable work load. Because of this, many groupings are combined as in Total Ciliates or Other Bluegreen Algae. The resolution of structure is therefore not as

detailed as in the SAM protocol. On the other hand it may be argued that the SAM method has less structure because of its lower number of species.

A list of our data analysis techniques that are used for both methods are listed in Table 6. The comparisons made here concentrate upon the NMCAA tool, but other methods are available. Again, the very different structures of the systems can affect the data analysis. The occurrence of numerous species in the MFC, many of them rare, can make conventional data analysis difficult since rare organisms may be absent in many of the sample collections.

Comparison of Patterns in the SAM and MFC Test Results

The two methodologies have quite contrasting means of introducing organisms to the systems, and the operational volumes and surface to volume ratios are quite different. One manifestation of these differences is likely in the erratic nature of the clustering of the MFC compared to the SAM experiments conducted with the same toxicant. In Figure 4a, the occurrence of significant clustering in regards to treatment group follows a distinctive pattern for the SAM experiment, an initial significant clustering followed by a convergence of the treatments and then a re-emergence of the clustering. The MFC experiment reflects a much noisier pattern, one that calls into question whether or not the observed significant clustering is an artifact. Figure 4b also demonstrates the noise inherent in the MFC as compared to the SAM system reflected in the NMCAA results.

In spite of the noise, and especially in the Jet-A experiments, an early and late period where the treatment groups are distinguishable seem to exist. In both sets of experimental protocols, Jet-A would have been seen to have generated more of an impact compared to JP-4, judging by the occurrence of significant clustering related to treatment effect.

As judged by the NMCAA results, none of the test systems demonstrated a recovery toward a stable system. This lack of recovery is reflected in both the significance of the clustering relative to treatment and

the changing in the important variable rankings over sampling days. As an example, compare the last 3 sampling days for the JP-4 MFC experiment. The only variable deemed as important on all three days is "optical density". "Other Bluegreens", "Ostracod 2", and "P/R" are found on two of the sampling dates. The variables pH, *P. bursaria* and Nitzschia are found on only one sampling date each. The rapidly changing significance values found in both MFC tests also indicate a dynamic and rapidly evolving system.

Generic Multispecies Toxicity Tests

Microcosm testing strategies provide a greater dimensionality to toxicity testing, and resolve impacts that can not be extrapolated from single species toxicity tests. The MFC and the SAM do not try to simulate specific natural ecosystems, but they do utilize organisms having distinct interspecific and intraspecific interrelations and responses typical of natural environments. These methods also display many of the structural and functional properties of ecosystems, e.g., photosynthetic production/respiration dynamics, competition and succession, grazing effects, and nutrient cycling [3,21,22]. Microbial process are present and degradation of xenobiotics and the potential impacts of degradation products can be studied [23].

The other main advantages of using these generic microcosms is that they are standardized in terms of species composition [3,21]. The importance of this simplicity and replicability in construction is that it allows closer examination of specific relationships and interactions in determining responses to direct and indirect effects, it reduces the dynamic heterogeneity that could potentially diffuse or hide effect responses, and it allows the comparison of results obtained in different laboratories [21].

The comparability and replicability of construction of a generic system is also a weakness. Since environmental heterogeneity, migration, colonization and other population, metapopulation and community level interactions are not modeled well in these systems, effects of toxicants upon these parameters will be difficult to ascertain. Numerous species representative of aquatic systems are not

included, for example fish and macrobenthos, and these organisms would be difficult of incorporate given the small size of the system.

Design Suggestions for Multispecies Toxicity Tests

The comparison of the two methods described here, along with our previous experience with microcosms and data analysis of these systems [16,17,18,23,24,25,26] leads us to suggest several improvements for the performance of multispecies toxicity tests. In several instances the suggestions are specific to the MFC and SAM systems, however many can be applied to systems regardless of size.

One of the most important aspects of any multispecies toxicity test is the realization by the investigators that these systems are models, inherently much more complex than computer simulations, of naturally occurring ecological systems. As has been demonstrated for the lakes studied by Katz et al. [27], the best predictor of the future behavior of a system is itself. All model ecosystems will be limited in their predictive power, however, a primary advantage of model systems is that they are likely to also include interactions, parameters and relationships that are currently unknown and therefore impossible to simulate in an explanation based system. Because of this fact, multispecies toxicity tests are powerful tools in the investigation and eventual understanding of toxicant impacts in naturally occurring systems. Our suggestions are made in this light.

Parameter Selection, Measurement and Sampling Frequency

In both microcosm protocols, the parameters measured and the analyses conducted, focus primarily on the biological structural components, including a few physical parameters, e.g., pH, dissolved oxygen, conductivity, and alkalinity. Species are identified and enumerated during the course of the experiment, to determine changes in diversity and abundance patterns. An important consideration is that these parameters are easily measured given the limited volumes and manpower requirements of performing the SAM or MFC tests. The premise of using this approach is that focusing on the functions, interactions, and responses of the individual parts will reveal ecosystem level dynamics [28]. Each population variable can

serve as an axis to track the movement of the system through ecosystem space. This approach is not without theoretical support. Ecosystems as perceived by the organisms are multidimensional. The Hutchinsonian idea of organisms and populations residing in a n-dimensional hypervolume is the basis of current niche theory [29]. The n-dimensional niche hypervolume is the ecosystem with all its components as perceived by the population. The variability of these parameters over time as well is used to account for the variety of species within the ecosystem [30,31,32].

Other parameters should also be sampled, if possible, to increase the resolution of the toxicity tests. There are limitations to the using of components to assess effects to the whole ecosystem. Microbial processes often dominate the metabolism of aquatic systems, yet procaryotic populations are difficult to measure and their rapid turnover times makes frequent sampling necessary. Since a 24 hr period can be as many as 48 generations in procaryote populations, sampling on the scale of hours would be necessary. Although the population structure of filter feeding organisms can give an indication of the procaryotic assemblage, other parameters can give a more direct indication of the status of the procaryotic community. Among these parameters are productivity/respiration ratios; total CO_2 efflux; biochemical rates; nutrient cycling; dissolved oxygen concentrations; pH; substrate decomposition rates; toxicant degradation rates; and accumulation rates of metabolic by-products [28,33].

Cross Inoculation

The purpose of cross inoculation among replicate systems is generally seen as a means of ensuring the homogeneity of the test systems prior to treatment. However, this principally sets each replicate as an island with frequent migration that will maintain each system with a larger number of species than normal for that particular island size. Species that would normally become extinct are re-supplied in the inoculum. Upon the elimination of the cross inoculation followed by the toxicant addition two factors are operating. First, a reduction in species as rare organisms become extinct. Second, the effects of the toxicant begin to operate. In effect, each of the 24 replicates starts from a different location in ecological space, no

control can be exercised to force them into similarity, and finally a toxicant impacts the system. Cross inoculation seems to unduly complicate the methodology without an increase in sensitivity.

Data Analysis

We strongly advocate the use of multivariate and Artificial Intelligence derived tools for the analysis of multispecies toxicity tests. A variety of methods have been developed that have the potential of revolutionizing our analysis of the dynamics of multispecies systems and their application to the risk assessment process has been discussed [34].

Normalized Ecosystem Strain (NES) has been developed by Kersting [8] as a means of describing the impacts of several materials to a three compartment microecosystems containing an autotrophic, herbivore and decomposer subsystems. These variables in the unperturbed control systems are used to calculate the normal operating range (NOR) of the microecosystem. The NOR is the 95 per cent confidence ellipsoid of the unperturbed state of a system. The center of the NOR is defined as the reference point for the calculation of the NES. The NES is calculated as the quotient of the Euclidean distance from a state to the reference state divided by the distance from the reference state to the 95 percent confidence (also called tolerance) ellipsoid, along the vector that connects the reference state to the newly defined state. A value of 1 or less indicates that the new state is within the 95 percent confidence ellipsoid, values greater than 1 indicate that the system is outside this confidence region. The sensitivity of the NES increased sensitivity as the number of variables used to describe the system increased.

Apparently as an independent development, A.R. Johnson [9] proposed the idea of using a multivariate approach to the analysis of multispecies toxicity tests. This state space analysis is based upon the common representation of complex and dynamic systems as an n-dimensional vector. A vector can be assigned to describe the motion of the system through this n-dimensional space to represent successional changes, evolutionary events, or anthropogenic stressors. The direction and position information form the trajectory of the state space and this can be plotted over time.

Another important application proposed by Johnson [10] was the use of multivariate analysis to identify diagnostic variables that can be applied in the monitoring of ecosystems. Diagnostic variables, if reliable in differentiating anthropogenically stressed systems from control systems would be extremely valuable in monitoring for compliance and in determining clean up standards.

We suggest the use both metric and nonmetric clustering for the evaluation of multispecies toxicity tests. In addition visualization tools such as the space-time worms aide in the evaluation of ecosystem dynamics. We have found these methods to lead to new insights into the nature of the dynamics of these tests systems. In the above sections we have summarized the methods. In the context of multispecies toxicity tests, the use of this approach has lead us to identify and compare the sensitivities of two microcosm methods and to arrive at the conclusion that the recovery of these systems is likely an illusion of perspective. These findings along with research along several fronts has lead us to conclude that the assumption of stability and recovery in ecological systems is erroneous [17].

Ecosystem Dynamics and the Importance of Nonequilibrium Conditions

The return of a system to its pre-existing state, structurally, metabolically and dynamically, is a classical definition of recovery. Stability confers upon a system with the ability to recover to a previous state. It has often been assumed the stability is a property of persistent ecological systems. It has even been suggested that the examination of stability and the measurements of resilience and recovery are the most appropriate attributes to be studied in multispecies toxicity tests [35]. Stability measurments are even advocated in spite of evidence indicating that such a property may not exist [36]. Even in situations where an equilibrium does not occur it is assumed that given more time that replicate systems will converge toward an equilibrium condition [37]. As comforting as an assumption of ecological stability may be, there is an increasing amount of data that indicate that stable systems may be the exception.

The return of a system to its pre-existing state, structurally, metabolically and dynamically, is a classical definition of recovery. In regard to populations, Connell and Sousa [38] examined a great deal of the literature on population dynamics and found stability as return to original conditions extremely rare.

Andrews [39] in a study of tropical lizards found that the population dynamics are unstable. Hypothesized causes are the rapid population turnover and the complexity of a food web. Over the last ten years it has in fact been found that many populations exhibit chaotic dynamics [40,41]. Although density dependent regulation is operating, the populations are characterized by large unpredictable fluctuations that are inherently unpredictable. In fact there is ample theory that predicts the inherent instability of large dynamic and connected systems [42,43]. Given the unpredictability of outcomes in a variety of theoretical [44] and experimental [45] cases, an assumption about the reality of stability and the reliance upon the measurement of recovery seems improper.

There is now a considerable amont of theory [46] that indicates systems that are too stable or frozen actually exhibit lower overall fitness and can lack the ability to respond to environmental alterations. While there may be order in ecological systems, it is unlikely that stability is the governing parameter. Perhaps it is time to discard a restrictive paradigm and search for new explanations of the dynamic behavoir of ecological systems perturbed by chemicals.

The studies compared here also provide several additional examples of systems that t are not driven toward an equilibrium. Indeed, given the similarity of patterns there may be more fundamental rules that govern the dynamics of systems under chemical stress. Indeed, the inherent dynamics of ecological systems and the study of their causes and outcomes should be the emphasis of multispecies toxicity testing.

Acknowledgments -- We would like to thank the Institute of Environmental Toxicology and Chemistry support staff for contributing to the production of this document, notably L. Keel, S. Rodgers, R. Sahakian, R. Sandberg, L.R. Cooke and N. Shough. L. Holmquist and T. Litwiller contributed to the editing. This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

REFERENCES

1. **Fisher, L.** 1992. Memorandum: Decisions on the Ecological, Fate and Effects Task Force. Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency.
2. **Gearing, J. N.** 1989. The roles of aquatic microcosms in ecotoxicologic research as illustrated by large marine systems. *Ecotoxicology: Problems and Approaches.* S.A. Levin, ed., Springer Verlag, pp. 411-470.
3. **Giddings, J. M.** 1983. Microcosms for assessment of chemical effects on the properties of aquatic ecosystems. In J. Saxena, ed, Hazard Assessment of Chemicals. Current Developments, Vol. 2. Academic Press, Inc., New York, pp. 45-94.
4. **Stay, F.S., T.E. Flum, L.J. Shannon and J.D. Yount.** 1989. An assessment of the precision and accuracy of SAM and MFC microcosms exposed to toxicants. In *Aquatic Toxicology and Hazard Assessment: 12th Volume, ASTM STP 1027,* U.M. Cowgill and L.R. Williams, eds., American Society for Testing and Materials, Philadelphia, PA, pp. 189-203.
5. **Conquest, L.L. and F.B. Taub.** 1989. Repeatability and reproducibility of the Standard Aquatic Microcosm: Statistical properties. In *Aquatic Toxicology and Hazard Assessment: 12th Volume, ASTM STP 1027,* U.M. Cowgill, L.R. Williams, eds., American Society for Testing and Materials, Philadelphia, PA, pp. 159-177.
6. **Kersting, K.** 1984. Development and use of an aquatic micro-ecosystem as a test system for toxic substances. Properties of an aquatic micro-ecosystem IV. *Int. Revue ges. Hydrobiol.* 69:567-607.
7. **Kersting, K.** 1985. Properties of an aquatic micro-ecosystem V. Ten years of observations of the prototype. *Verh. Internat. Verein. Limnol.* 22:3040-3045.
8. **Kersting, K.** 1988. Normalized ecosystem strain in micro-ecosystems using different sets of state variables. *Verh. Internat. Verein. Limnol.* 23:1641-1646.

9. **Johnson, A.R.** 1988a. Evaluating ecosystem response to toxicant stress: a state space approach. In *Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP* 971, W.J. Adams, G.A. Chapman and W.G. Landis, eds., American Society for Testing and Materials, Philadelphia, PA, pp. 275-285.

10. **Johnson, A.R.** 1988b. Diagnostic variables as predictors of ecological risk. *Environmental Management* 12:515-523.

11. **Matthews, G.B. and J. Hearne.** 1991. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(2):175-184.

12. **Taub, F.B.** 1989. Standardized aquatic microcosms. *Environm. Sci. Technol.* 23:1064-1066.

13. **ASTM E 1366-91.** 1991. *Standard Practice for the standardized aquatic microcosm: fresh water*, Vol 11.04, pp 1017-1051, American Society for Testing and Materials, Philadelphia, PA.

14. **Leffler, J. W.** 1984. The use of self-selected, generic aquatic microcosms for pollution effects assessment. Concepts in Marine Pollution Measurements. H. H. White, ed. University of Maryland, College Park. pp. 139-157.

15. **Shannon, L.J., T.E. Flum, R.L. Anderson and J.D. Yount.** 1989. Adaptation of mixed flask culture microcosms for testing the survival and effects of introduced microorganisms. In *Aquatic Toxicology and Hazard Assessment: 12th Volume, ASTM STP 1027*, U.M. Cowgill and L.R. Williams, eds., American Society for Testing and Materials, Philadelphia, PA, pp. 224-239.

16. **Landis, W.G., R.A. Matthews, A.J. Markiewicz, N.A. Slough and G.B. Matthews.** 1993a. Multivariate analyses of the impacts of the turbine fuel Jet-A using a microcosm toxicity test. *J. Environ. Sci.* 2:113-130.

17. **Landis, W.G., R.A. Matthews, A.J. Markiewicz and G.B. Matthews.** 1993b. Multivariate analysis of the impacts of the turbine fuel JP-4 in a microcosm toxicity test with implications for the evaluation of ecosystem dynamics and risk assessment. *Ecotoxicology* 2:271-300.

18. **Matthews, G.B., R.A. Matthews and W.G. Landis.** *In press.* Nonmetric clustering and association analysis: Implications for the evaluation of multispecies toxicity tests and field monitoring. In *Environmental Toxicology and Risk Assessment: Third Volume, ASTM 1218,* J.S. Hughes, G.R. Biddinger and E. Mones, eds., American Society for Testing and Materials, Philadelphia, PA, 1994.

19. **Press, W.H., B.P. Flannery, A.A. Teukolsky and W.T. Vetterline.** 1990. *Numerical Recipes in C, the Art of Scientific Computing.* Cambridge University Press, New York, NY.

20. **Marklewicz, A.J.** 1994. Comparison of the Degradation of Water Soluble Components in Jet Fuel Using the Standard Aquatic Microcosm (SAM) and Mixed Flask Culture (MFC). M.S. Thesis, Western Washington University.

21. **Suter, G.W. II.** 1993. A critique of ecosystem health concepts and indexes. *Environ. Toxicol. Chem.* 12:1533-1539.

22. **Taub, F.B.** 1984. Synthetic microcosms as biological models of algal communities. In L. E. shubert, ed., Algae as Ecological Indicators. Academic Press, Inc. London pp. 363-394.

23. **Landis, W.G., M.V. Haley and N.A. Chester.** 1993c. The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In *Environmental Toxicology and Risk Assessment ASTM STP -1167,* W.G. Landis, J. Hughes and M. Lewis, eds., American Society for Testing and Materials, Philadelphia, PA, pp. 159-177.

24. **Haley, M.V., D.W. Johnson and W.G. Landis.** 1988. The aquatic toxicity of brass dust. In *Aquatic Toxicology and Environmental Fate: Tenth Volume ASTM STP -971,* W. Adams, G. Chapman and W.G. Landis, eds., American Society for Testing and Materials, Philadelphia, PA, pp. 468-479.

25. **Matthews, G. and R. Matthews.** 1990a. A model for describing community change. In Pesticides in Natural Systems: How Can Their Effects Be Monitored? Proceedings of the Conference, Environmental Research Laboratory/ORD, Corvallis, OR. EPA 9109/9-91-011.

26. **Sandberg, R.** 1993. Investigation of the Effects of a Pulsed Release of Jet-A Turbine Fuel From Sediments Using a Modified Mixed Flask Culture (MFC) Microcosm. M.S. Thesis, Western Washington University.

27. **Katz, T.K., T.M. Frost and J.J. Magnuson.** 1987. Inferences from spatial and temporal variability in ecosystems: Long-term zooplankton data from lakes. *Amer. Nat.* 129:830-846.

28. **O'Neill, R. V. and J. B. Waide.** 1981. Ecosystem theory and the unexpected: Implications for environmental toxicology. In B. W. Comby, ed., Management of Toxic Substances in Our Ecosystems. Taming the Medusa. Ann Arbor Science Publishers, Inc. pp. 43-73.

29. **Hutchinson, G.E.** 1959. Concluding remarks. *Cold Spring Harbor Symposium on Quanitative Biology.* 22:415-427.

30. **Hutchinson, G.E.** 1961. The paradox of the plankton. *Amer. Nat.* 95:137-143.

31. **Richerson, P., R. Armstrong** and **C.R. Goldman.** 1970. Contemporaneous disequilibrium, a new hypothesis to explain the "paradox of the plankton". *Proc. Natl. Acad. Sci. U.S.A.* 67:1710-1714.

32. **Tilman, D.** 1982. *Resource Competition and Community Structure.* Princeton University Press, Princeton, NJ, pp. 11-138.

33. **Sugiura, K.** 1992. A multispecies laboratory microcosm for screening ecotoxicological impacts of chemicals. *Env. Tox. Chem.* 11:1217-1226.

34. **Landis, W.G., G.B. Matthews, R.A. Matthews and A. Sergeant.** *In press 1994.* Application of multivariate techniques to endpoint determination, selection and evaluation in ecological risk assessment. *Environ. Toxicol. Chem.*

35. **Johnson, M. L. D. G. Huggins and F. deNoyelles, Jr.** 1991. Ecosystem modelling with LISREL: a new approach for measuring direct and indirect effects in ecosystem level ecotoxicological testing. *Ecological Applications* 1:383-398.

36. **Dewey, S. L. and and F. deNoyelles, Jr.** 1994. On the use of ecosystem stability measurements in ecological effects testing. In Aquatic Mesocosm Studies in Ecological Risk Assessment. Eds R. L. Graney, J. H. Kennedy, J. H. Rodgers. Lewis Publishers, Boca Raton., pp. 605-625.

37. **Rosenzweig, M.S. and A.L. Bulkema, Jr.** 1994. Phytoplankton colonization and seasonal succession in new experimental ponds. *Environ. Toxicol. Chem.* 13:599-605.

38. **Connell, J.H. and W.P. Sousa.** 1983. On the evidence needed to judge ecological stability or persistence. *Am. Nat.* 121:789-824.

39. **Andrews, R.M.** 1991. The population stability of a tropical lizard. *Ecology* 72:1204-1217.

40. **Schaffer, W.M. and M. Kot.** 1985. Do strange attractors govern ecological systems? *Bioscience* 35: 342-350.

41. **Hastings, A., C.L. Hom, S. Ellner, P. Turchin and H.C. Godfray.** 1993. Chaos in ecology: Is mother nature a strange attractor? *Annu. Rev. Ecol. Syst.* 24:1-33.

42. **Gardner, M. R. and W. R. Ashby.** 1970. Connectance of large dynamic (cybernetic) systems: critical values for stability. *Nature* 228:784.

43. **May, R.M.** 1972. Will a large complex system be stable? *Nature*, Vol. 238, pp. 413-414.

44. **Yodzis, P. 1988.** The indeterminacy of ecological interactions as perceived through perturbation experiments. *Ecology* 69:508-515.

45. **Tilman, D. and and D. Wedin.** 1991. Oscillations and chaos in the dynam ics of a perinnial grass. *Nature* 353: 653-655.

46. **Kauffman, S. A.** 1993. *Origins of Order: Self Organization and Selection in Evolution.* Oxford University Press, New York pp 29-279.

Table 1. Comparison of ranked variables for the SAM experiments.

Day	Important Variables in Determining Clusters in Rank Order
Jet A Standardized Aquatic Microcosm	
11	M. Daphnia, Chlorella, Chlamydamonas, Ulothrix, S. Daphnia,Selanastrum,Scenedesmus
14	S. Daphnia, M. Daphnia-Selenastrum[1], Chlamydamonas, Chlorella, L. Daphnia, Ankistrodesmus
18	Ankistrodesmus, S. Daphnia, Chlorella, Chlamydamonas, Selanstrum, L. Daphnia
21	Ankistrodesmus, S. Daphnia, L. Daphnia-M. Daphnia, Scenedesmus
25	Scenedesmus, S. Daphnia, L. Daphnia, Chlorella, Philodina-M. Daphnia
28	Ankistrodesmus, L. Daphnia, Scenedesmus
32	S. Daphnia, M. Daphnia, Ankistrodesmus, Chlorella
35	Ankistrodesmus
39	M. Daphnia-Selenastrum, **Ostracod**-Ankistrodesmus
42	M. Daphnia, **Ostracod**, Scenedesmus
46	Scenedesmus, Ankistrodesmus, S. Daphnia. M. Daphnia
49	Chlorella, Philodina, Ankistrodesmus, Lyngbya
53	Ankistrodesmus, **Ostracod**, Chlorella
56	M. Daphnia-Scenedesmus, Ankistrodesmus, Lyngbya
60	Lyngbya, M. Daphnia, Philodina, Chlorella
63	Chlorella, Ankistrodesmus, Philodina, **Ostracod**
JP-4 Standardized Aquatic Microcosm	
11	**Selanastrum**, M. Daphnia, Chlorella, Ankistrodesmus
14	Selenastrum, S. Daphnia, M. Daphnia-Ankistrodesmus[1], L. Daphnia-Stigeoclonium
18	Scenedesmus, Selanstrum, Ankistrodesmus, S. Daphnia, Chlorella, L. Daphnia
21	Scenedesmus, Ankistrodesmus, Chlamydomonas
25	Chlorella, S. Daphnia

28	**Chlorella, Ankistrodesmus-Lyngbya, Philodina**
32	**Ostracod**
35	**Ostracod, Philodina, Scenedesmus**
39	Scenedesmus, S. Daphnia
42	Lyngbya, S. Daphnia, Philodina, Ankistrodesmus
46	M. Daphnia
49	Scenedesmus, Chlorella, Philodina
53	Chlorella, Philodina
56	M. Daphnia-S. Daphnia
60	S. Daphnia, **Ostracod**, Lyngbya
63	Chlorella, S. Daphnia, M. Daphnia, Lyngbya

[1] Hyphen between variables denotes equal rank.

Table 2. Comparison of ranked variables for the MFC experiments.

Day	Important Variables in Determining Clusters in Rank Order
Jet-A Mixed Flask Culture	
00	Ciliates-Flagellates, Optical Density, P/R
04	**Ostracod 2**, Other Ugreen-Flagellates[1], Chlorella, Amphipods
07	**Ostracod 2**, Other Diatoms-Optical Density, Scenedesmus, Chlorella, Other Ugreen
11	Other Diatoms, Nitzschia, pH, **Ostracod 1**, Other Bluegreen
14	Ankistrodesmus-Other Diatoms, Scenedesmus-Other Fgreen, Other Ugreen, Nitzschia
18	Other Diatoms, Nitzschia, Amphipods, Scenedesmus-Anabaena
21	Other Diatoms, Other Ugreen, Nitzschia
25	Other Ugreen, Other Diatoms, Optical Density, Anabaena
28	P/R, Ankistrodesmus, pH
32	Chlorella, Other Ugreen, Ankistrodesmus
35	Scenedesmus, Amphipods, Flagellates
39	Selenastrum, Scenedesmus-Amphipods, *Chlorella*, *Other Diatoms*, **Ostracod 1**
42	Other Ugreen, Nitzschia
46	**Ostracod 1**, Nitzschia-Other Bluegreen, Ciliates, *P. bursaria*
49	**Ostracod 1**-Other Diatoms, P/R, Chlorella, Flagellates
56	Chlorella, **Ostracod 1**, Ankistrodesmus-Flagellates
60	Chlorella, Amphipods, Other Diatoms
63	Chlorella, **Ostracod 2**, Nitzschia
67	Other Diatoms-**Ostracod 1**-pH, Ankistrodesmus-Nitzschia-**Ostracod 2**-Ciliates, *P. bursaria*
70	pH, Other Ugreen-Amphipods-Ciliates, P/R, **Ostracod 1**, Ankistrod
74	**Ostracod 2**, **Ostracod 1**, Ankistrodesmus-Other Bluegreen, Other Ugreen-P/R
77	pH, Scenedesmus-P/R, Other Fgreen-Flagellates-Optical Density, Chlorella

JP-4 Mixed Flask Culture

00 Selenastrum, Optical Density, P. bursaria, P/R, Scenedesmus

04 M. Daphnia, S. Daphnia, Chlorella, Other Bluegreen

07 Other Bluegreen, Chlorella, Ostracod 2, Lyngbya, Other Diatoms-Optical Density[1]

11 Other Bluegreen, M. Daphnia, Ciliates, S. Daphnia, Flagellates

14 M. Daphnia, Other Bluegreen, Scenedesmus, Flagellates-pH, S. Daphnia-Ciliates

18 pH, Lyngbya, M. Daphnia, Optical Density, Scenedesmus-Other Bluegreen, Rotifers, Ciliates

22 Rotifers, Selenastrum-Ciliates, Other Bluegreen, pH, Lyngbya

25 Other Fgreen, Ostracod 2, Scenedesmus-Ciliates-pH

29 Nitzschia, Other Diatoms, Copepod, P/R

32 Other Bluegreen, *P. bursaria*, pH

35 Ciliates-Optical Density, Other Bluegreen

39 pH, Other Ugreen, P/R

42 Optical Density, Other Diatoms, Selenastrum, Ciliates, Other Bluegreen

46 Flagellates, Ankistrod, *P. bursaria*, Ostracod 2, pH

49 Optical Density, Other Bluegreen, *P. bursaria*

53 Optical Density, Ostracod 2, P/R

56 Optical Density, Ostracod 2, Other Bluegreen, P/R, pH, Nitzschia

[1] Hyphen between variables denotes equal rank

Table 3. Comparison of the experimental designs of the SAM and MFC multispecies toxicity tests. The numbers of groups and replicates are identical in each system.

Experimental Design

Standardized Aquatic Microcosm		Mixed Flask Culture	
Number of groups:	4	Number of groups:	4
Number of replicates:	6	Number of replicates :	6
Reinoculation:	Once per week add one drop (circa 0.05 mL) to each microcosm from a mix of the ten species = 5 x 10 2 cells of each alga added per microcosm	Reinoculation:	Only reinoculated and cross innoculated during the maturation period.
Addition of test materials:	Add material on Day 7	Sampling frequency:	2 times each week
Sampling frequency:	2 times each week	Test duration:	6-8 weeks Allow to mature 6 weeks prior to treatment; track 6 to 8 weeks after exposure. Microcosms are rotated once a week in the environmental chamber during the experiment.
Test duration:	63 days		

Table 4. Comparisons of the physical and chemical structure of the SAM and MFC multispecies toxicity tests. The media are identical except for the addition of $NaHCO_3$ in the MFC protocol. Due to the reduced volume of the MFC and its container, the MFC has 1.5 times the surface to volume ratio of the SAM experimental unit.

Size, Medium and Sediment

Standardized Aquatic Microcosm	Mixed Flask Culture
One-gallon (3.8 L) glass jars are recommended; soft glass is satisfactory if new containers are used; measurements should be 16.0 cm wide at the shoulder, 25 cm tall with 10.6 cm openings.	1 L beakers covered with a large petri dish
Microcosm medium: 3 L T82MV	Microcosm medium: 900 mL of T82MV supplemented with 15 µg $NaHCO_3$ as an additional carbon source. into which 50 mL of inoculum was introduced
Sediment: Composed of silica sand (200 g), ground, crude chitin (0.5g), and cellulose powder (0.5 g) added to each container.	Sediment: 50 mL of acid washed sand

Table 5. Comparisons of the measurement endpoints of the SAM and MFC multispecies toxicity tests. Essentially the same levels of biological organization are included in both methods. In the calculation of clusters, derived variables are not particularly useful since they disproportionately weight certain measurements.

Measurement Endpoints

Standardized Aquatic Microcosm	Mixed Flask Culture
Primary Variables	**Primary Variables**
Population densities of inoculated organisms	Population densities of introduced organisms (often by classes such as diatoms, bluegreen bacteria, ostracods, protozoa etc.)
pH	pH
Photosynthesis/Respiration ratio	Photosynthesis/Respiration ratio
Optical Density	Optical Density
Analytical Chemistry of toxicant	Analytical Chemistry of toxicant
Nutrients	Nutrients
Bacterial counts	Bacterial counts
Derived variables	**Derived variables**
Algal Diversity	Algal Diversity
Total Algae	Total Algae
Available Algae	Available Algae
Total Daphnia	Total Daphnia
Total Invertebrates	Total Invertebrates

Table 6. Data analysis of the SAM and MFC multispecies toxicity tests. In our analyses, each system is analyzed using the same suite of statistical and artificial intelligence tools.

Data Analysis

Standardized Aquatic Microcosm	Mixed Flask Culture
Graphical Analysis (Plot the data) **Intervals of Non-significant Differences (IND)** **Metric Multivariate Statistics** **Non-metric multivariate Statistics-Riffle** **Projections-Space-time worms**	
Fewer species allow better identification and understanding of the potential role of each in the observed dynamics	Thousands of species and counting is often done at a variety of taxonomic levels. Not as much information on each of the organisms makes it difficult to assign roles and understand interactions. Many rare species

Figures

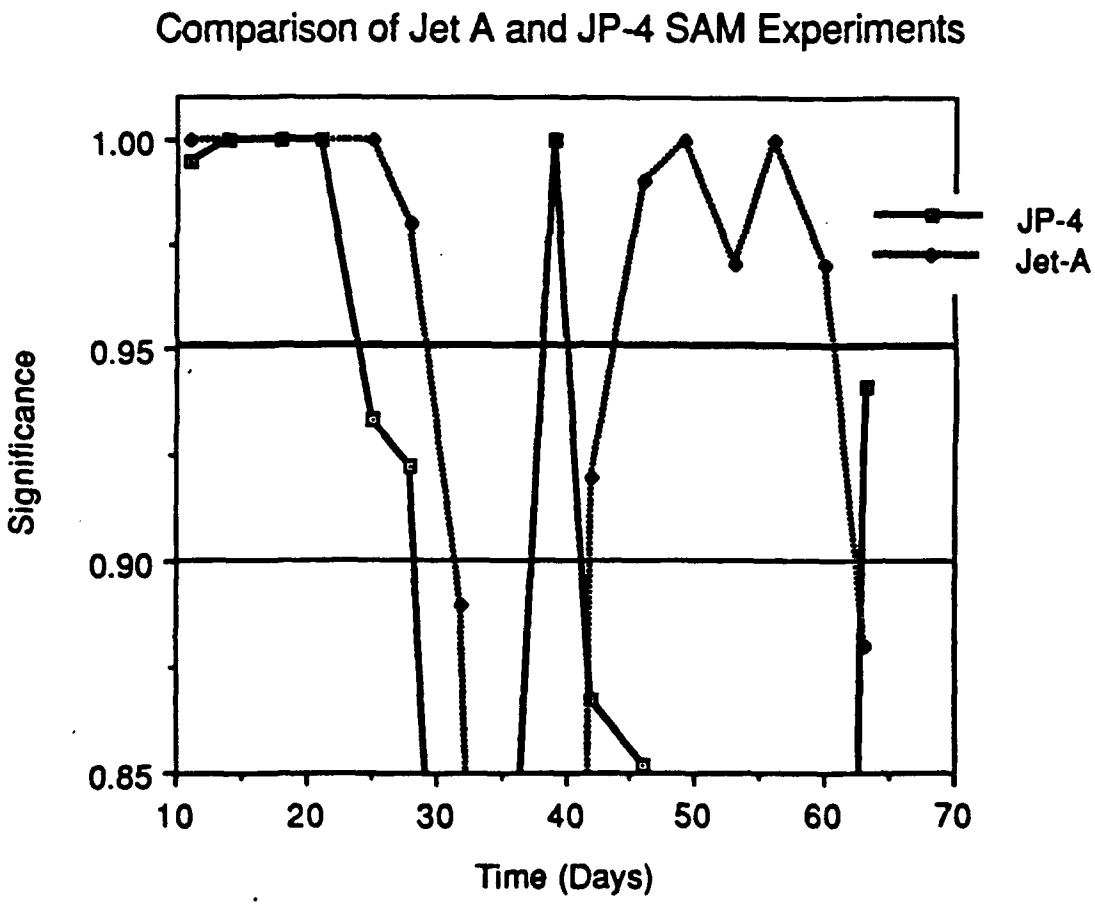
Figure 1. Comparison of the nonmetric clustering and association results for the Jet-A and JP-4 SAM.

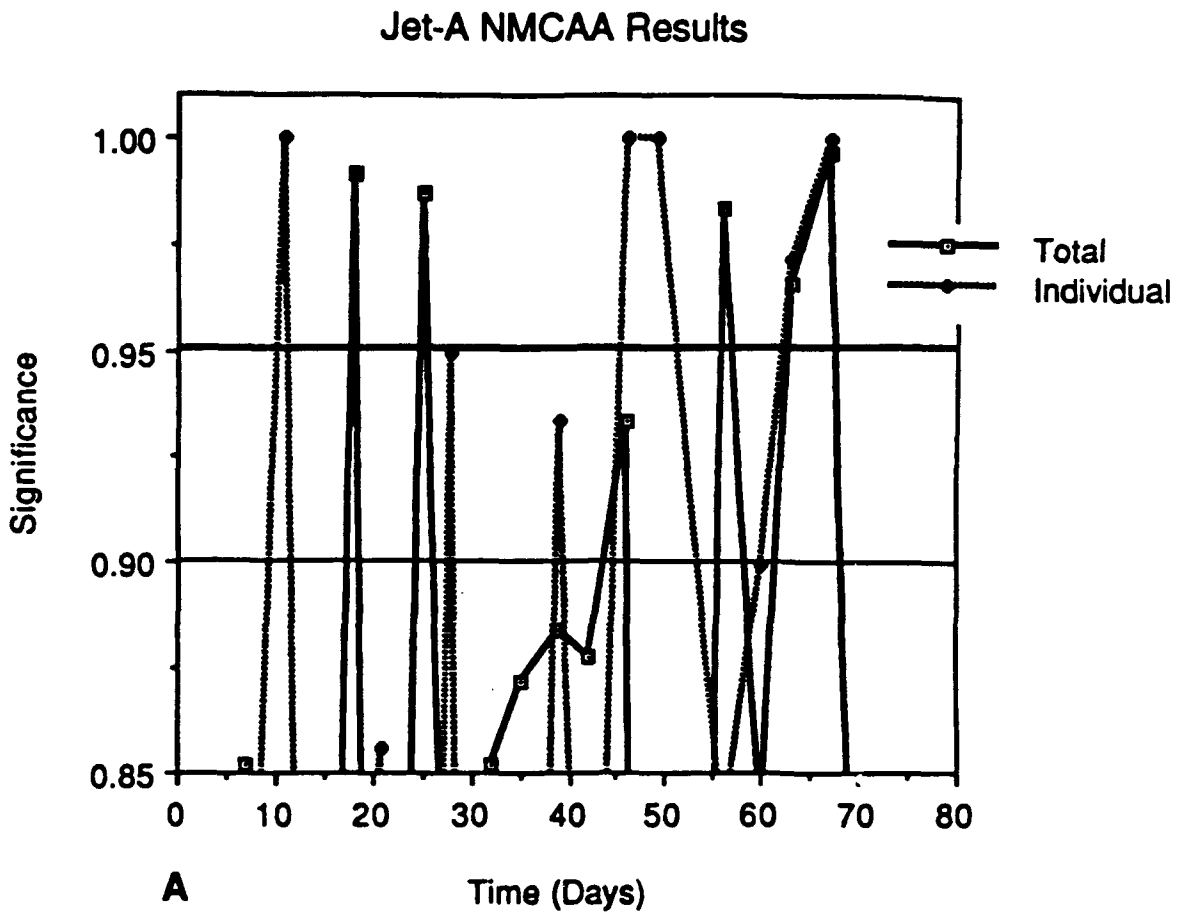
Figure 2. The use of individual based and group based data in the determination of nonmetric clustering and association results of the MFC experiments.

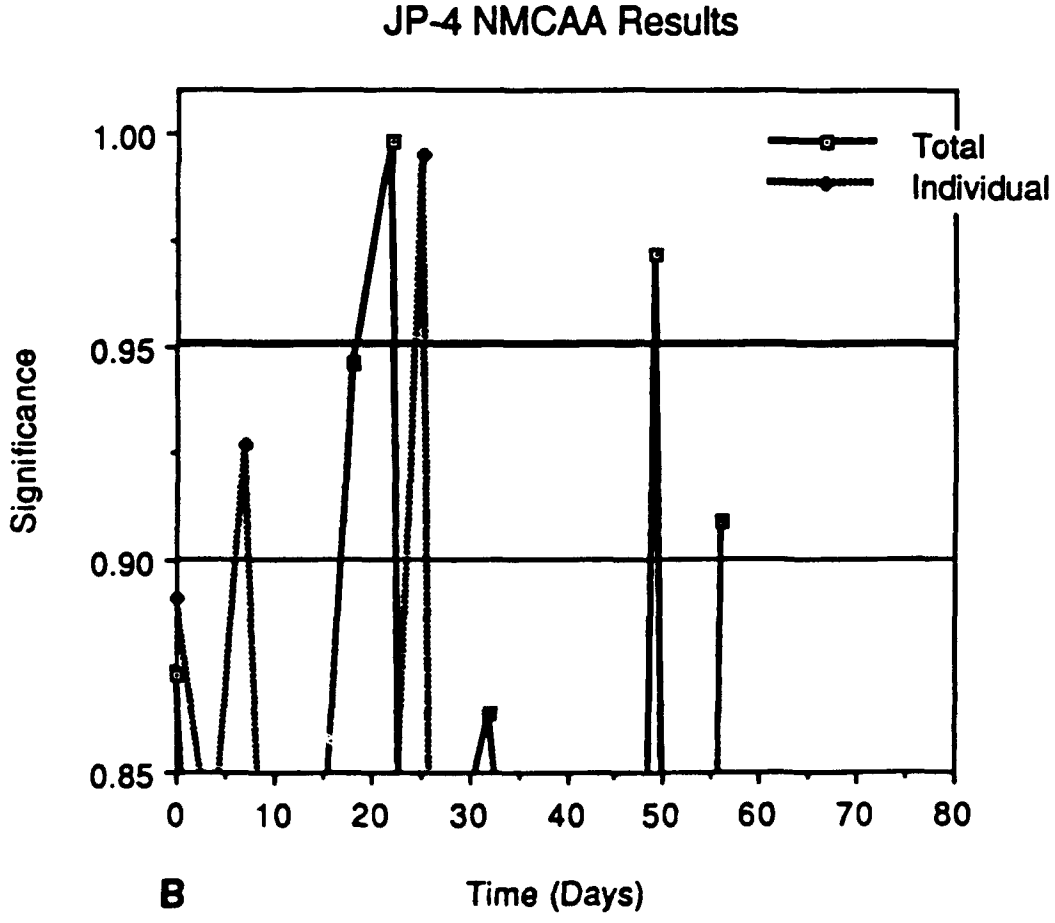
Figure 3. Comparison of the impact of Jet-A and JP-4 in the MFC using both individual and group based nonmetric clustering and association results.

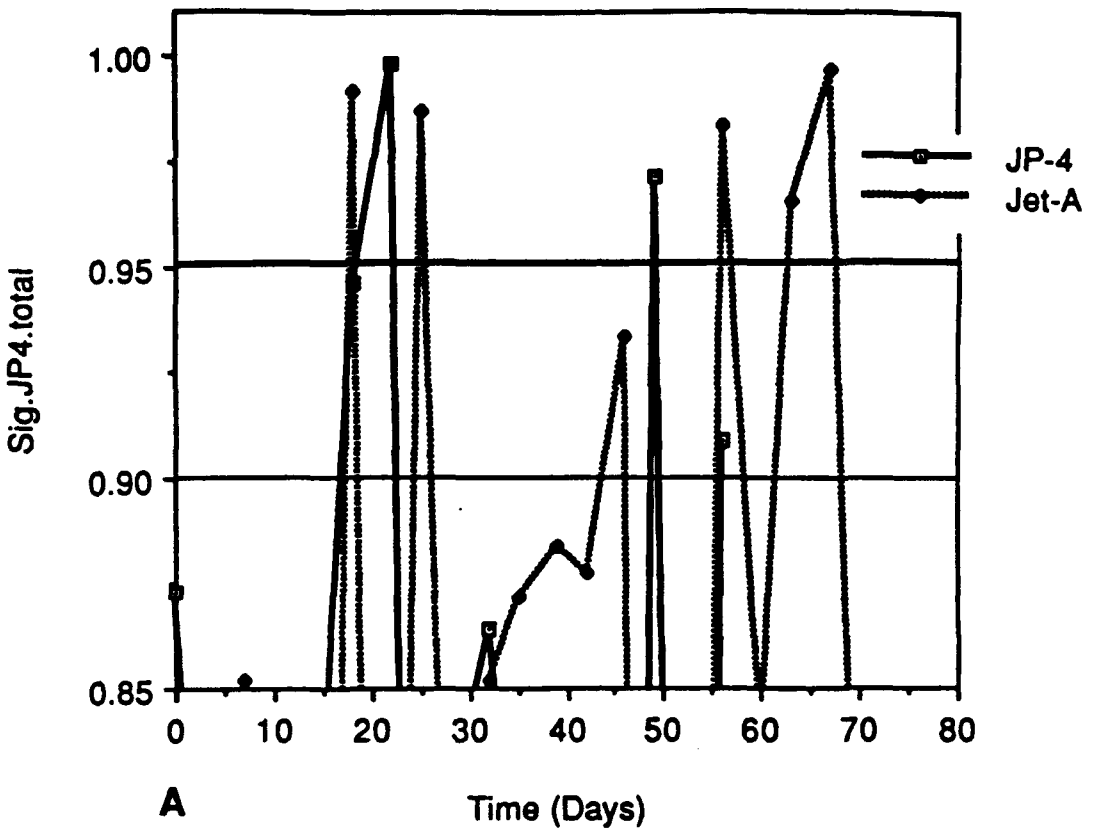
Figure 4. Comparison of the MFC and SAM nonmetric clustering and association results with Jet-A and JP-4.

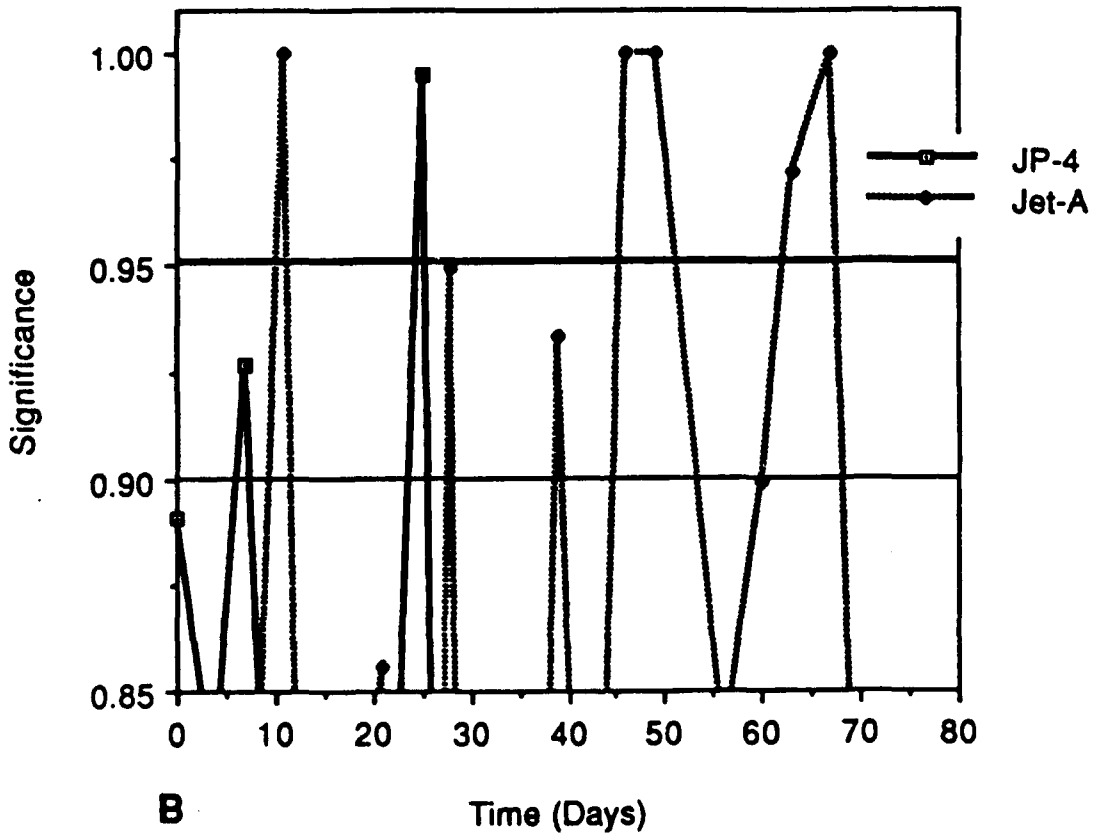
Figure 5. Comparison of the space-time worm projections for the Jet-A SAM and MFC experiments. In these projections time runs left to right and the microcosm axes are *Ankistrodesmus* and *Ostracods*. The divergence-convergence-divergence pattern appears in both experimental systems. Figure 5A portrays the Jet-A SAM system dynamics. There is a clear initial divergence followed by a convergence and towards the end of the experiment another divergence. Figure 5B uses the same projection to portray the dynamics of the Jet-A MFC. In the early part of the experiment the systems largely run parallel to each other. At the midway point large divergences beginthat appear more erratic and violent compared to the SAM system.

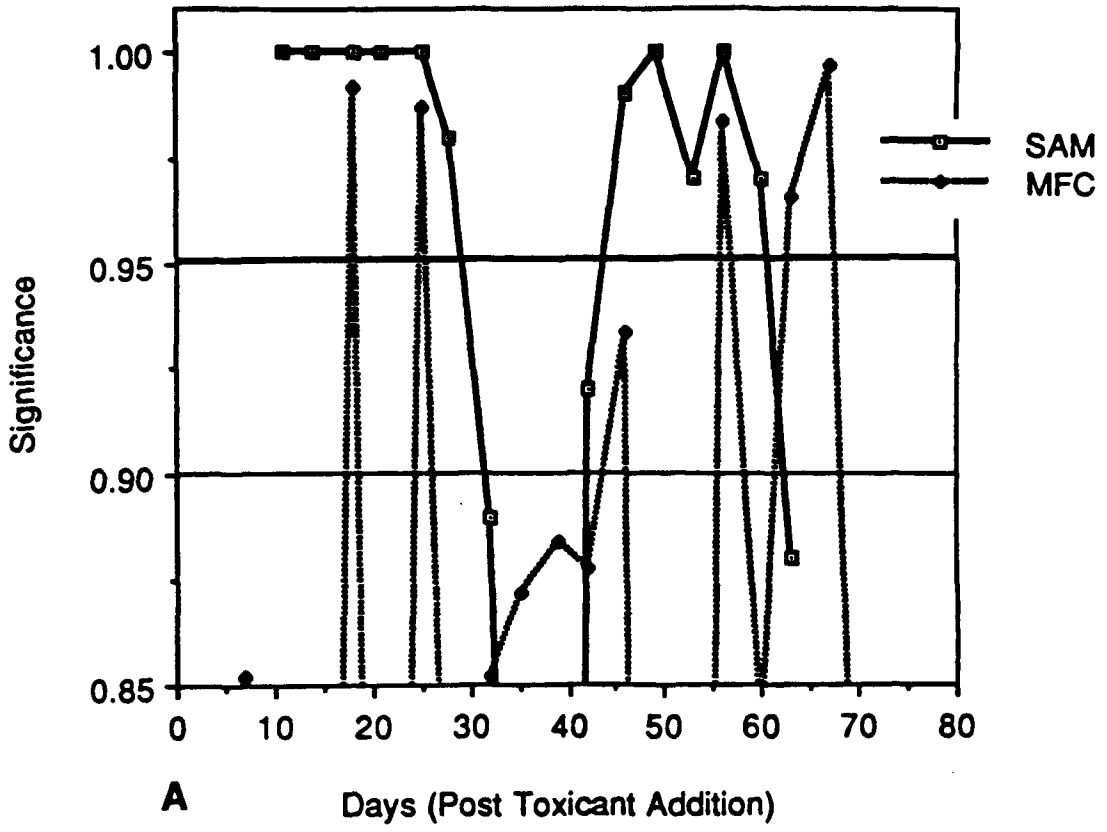

Comparison of Jet A and JP-4 SAM Experiments
1.00
0.95
0.90
0.85
Significance
10
20
30
40
50
60
70
Time (Days)
JP-4
Jet-A

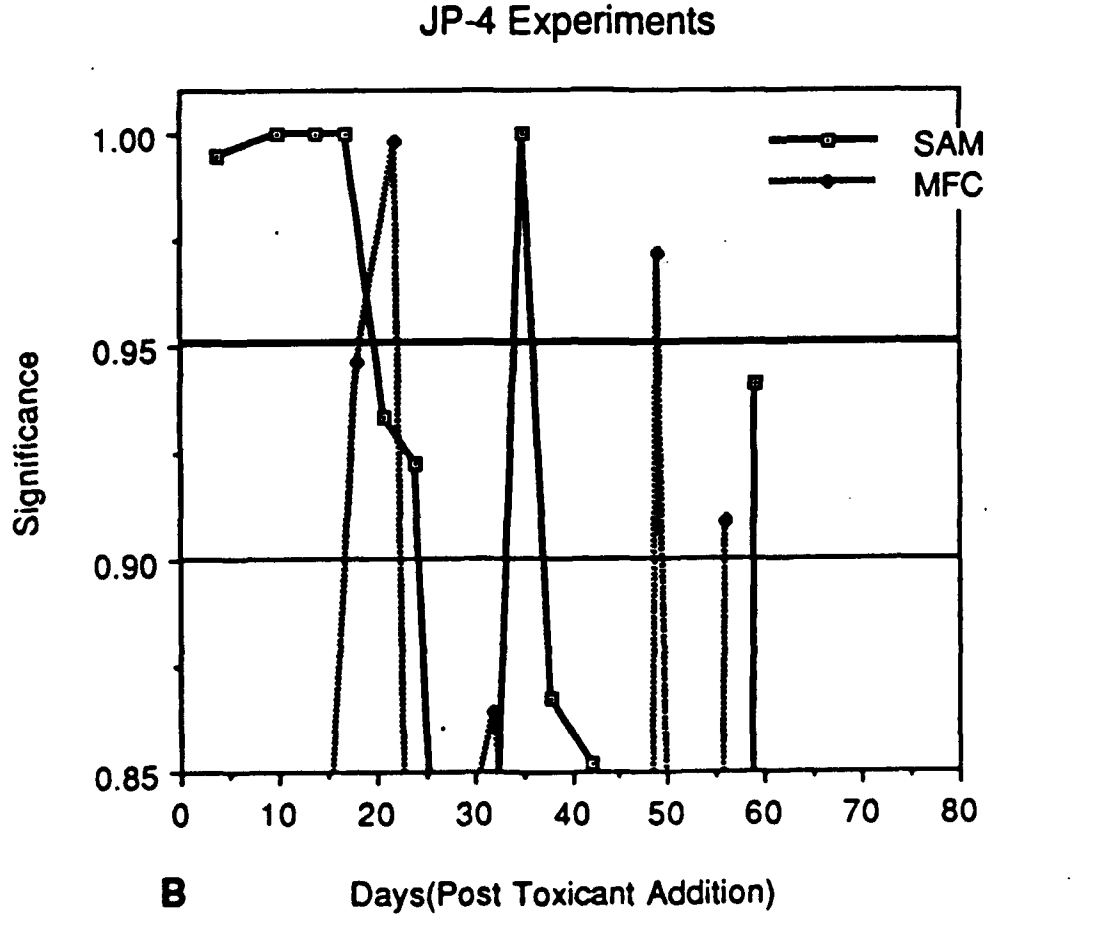

Jet-A NMCAA Results
1.00
0.95
0.90
0.85
Significance
0
10
20
30
40
50
60
70
80
Time (Days)
Total
Individual
A

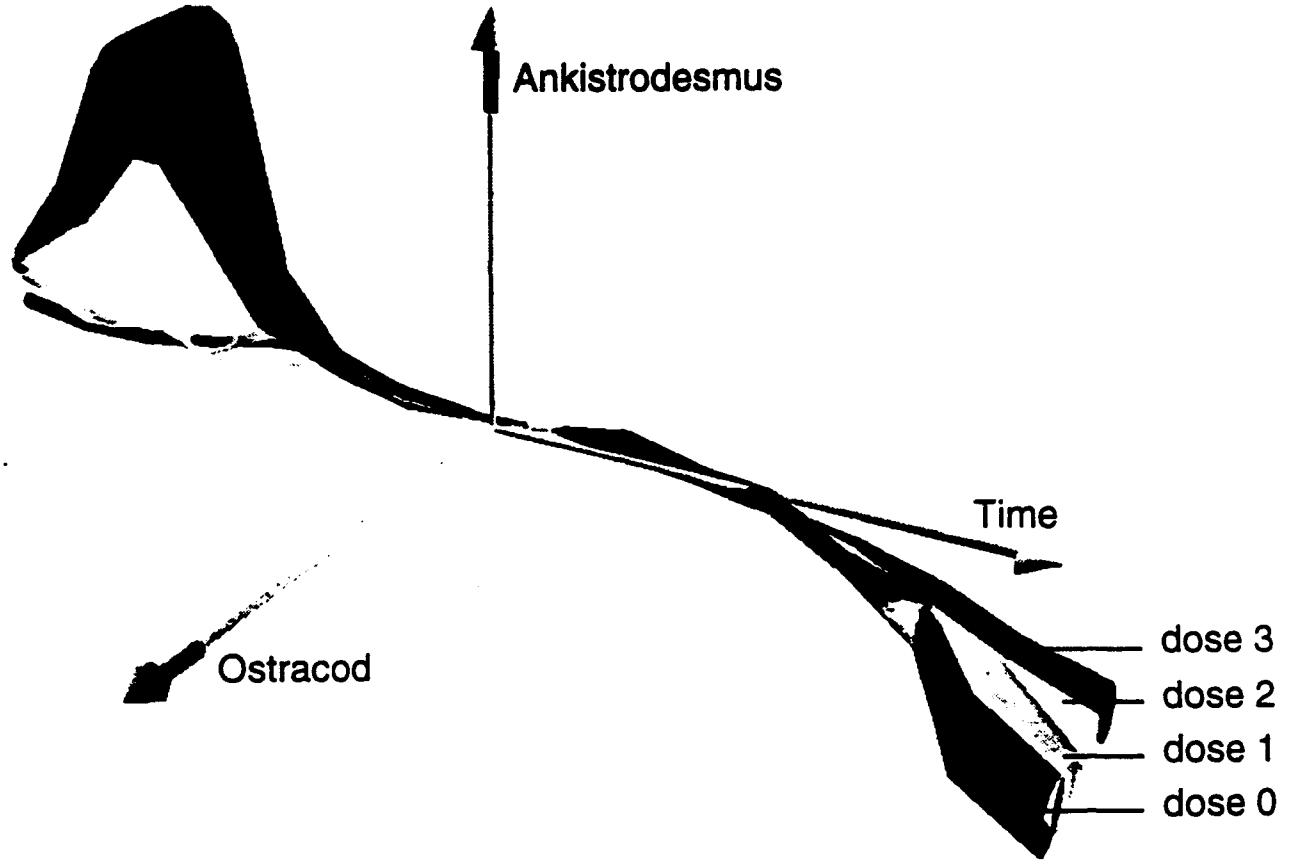

JP-4 NMCAA Results
1.00
0.95
0.90
0.85
Significance
0
10
20
30
40
50
60
70
80
Time (Days)
Total
Individual
B

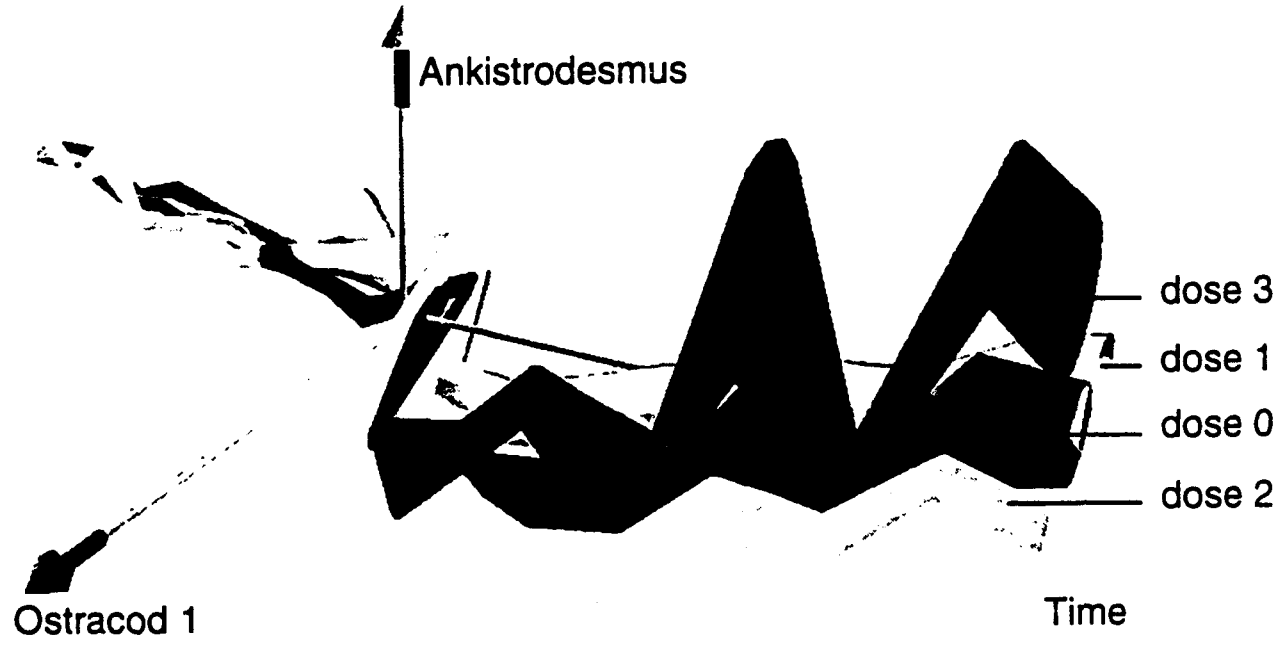

NMCAA Comparison using Total Counts
Sig.JP4.total
1.00
0.95
0.90
0.85
0
10
20
30
40
50
60
70
80
JP-4
Jet-A
A
Time (Days)


NMCAA Comparisons using Individual Counts
Significance
1.00
0.95
0.90
0.85
0
10
20
30
40
50
60
70
80
JP-4
Jet-A
B
Time (Days)


Jet A Experiments
1.00
0.95
0.90
0.85
Significance
0
10
20
30
40
50
60
70
80
SAM
MFC
A
Days (Post Toxicant Addition)


JP-4 Experiments
1.00
0.95
0.90
0.85
Significance
0
10
20
30
40
50
60
70
80
SAM
MFC
B
Days(Post Toxicant Addition)


Ankistrodesmus
Time
Ostracod
dose 3
dose 2
dose 1
dose 0


Ankistrodesmus
dose 3
dose 1
dose 0
dose 2
Ostracod 1
Time

APPENDIX B

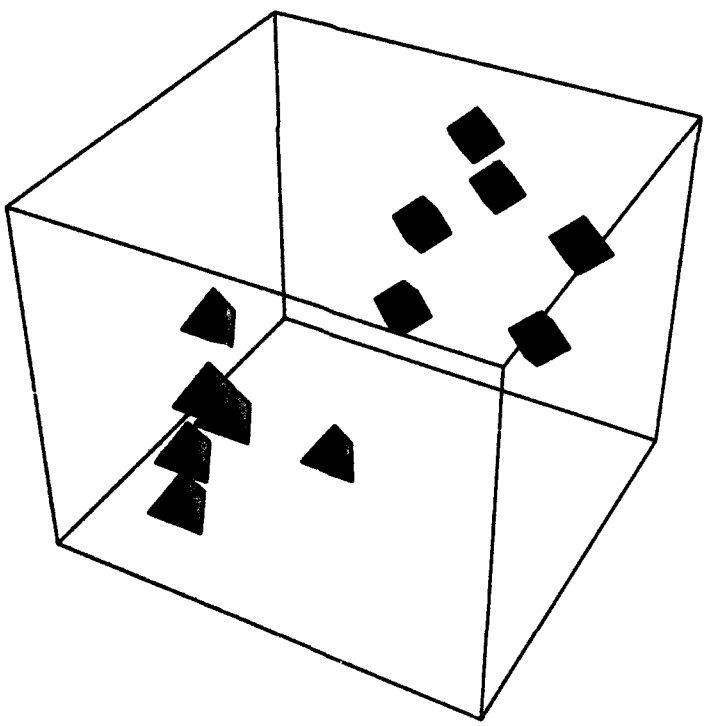
Short Course and
Mini Course Manuals

Nonmetric Clustering and Association Analysis in Ecotoxicology

SETAC 93 Short Course

Geoffrey Matthews
Mike Roze
Robin Matthews
Wayne Landis

November 14, 1993



Contents

1 Introduction

The goal of data analysis is the discovery of a model which fits the data. Statistical tools to accomplish this goal can differ in two ways: First, analysis tools differ in the kind of model which they fit to the data. For example, regression attempts to fit a linear subspace to the data points. Ordination attempts to fit a linear order to the data points. Clustering attempts to fit the data with a finite number of clusters, or subpopulations, each with distinct properties. More ambitious, lumped models, such as hydrologic models, attempt to fit the data with a model which mimics its causes. We call this choice of model for an analytic tool its *model bias*. Second, analysis tools differ in the criteria used for goodness of fit. Regression typically seeks to minimize the sum of the squared distances of the data points from the regression subspace, but other measures, such as the sum of absolute values, or the median distance, could be used. Similar measures are used in lumped models, where the model output is "calibrated" to match the available data by minimizing the distances. In clustering, the fitness criterion is usually the minimization of intra-cluster distance and simultaneous maximization of inter-cluster distance. The bias of the clustering procedure will then depend on the distance function or metric used. We call this aspect of an analysis tools its *fitness bias*.

2 Fitness Bias in Regression

Fitting curves to data is an important part of understanding the data. LC50 analysis, such as probit, tries to fit a sigmoid curve (such as $(1 + e^{-x})^{-1}$) to the mortality data. As an illustration of the importance of fitness bias in a simple task like curve fitting, consider the problem of fitting a curve to the data in Figure 1. The problem of model bias is usually well recognized. A linear fit, as in Figure 2, or a quadratic fit, as in Figure 3, are clearly inappropriate for the data in hand. Plots of the residuals would easily reveal curves that indicate the true nature of the trend has not been captured. An exponential fit, as in Figure 4, is clearly a better model, and the residuals and sums of squares would confirm this.

Our first hint of trouble arises, however, when we replot the data and the fitted curve on a log scale. The data on a log scale is shown in Figure 5, and the fitted exponential line is shown in Figure 6. The fit looks quite poor, and we could draw a much better line by hand. What went wrong?

To figure this out, note that the exponential line fits better on the right end than on the left end. Its fit is quite poor for small values of y. That is the clue to what went wrong. The "fitness" we sought was a "least squares" fit. It's obvious that the square of a big number is bigger than the square of a small number, and so our regression tried to get "closer" to big numbers than it did to small numbers.

Clearly this is inappropriate. In almost all data, the variance of samples is not independent of the mean; usually it is approximately proportional to the mean. The bigger the mean, the bigger the variance. Thus, we often say, "plus or minus five percent" while we rarely get a chance to say "plus or minus five."

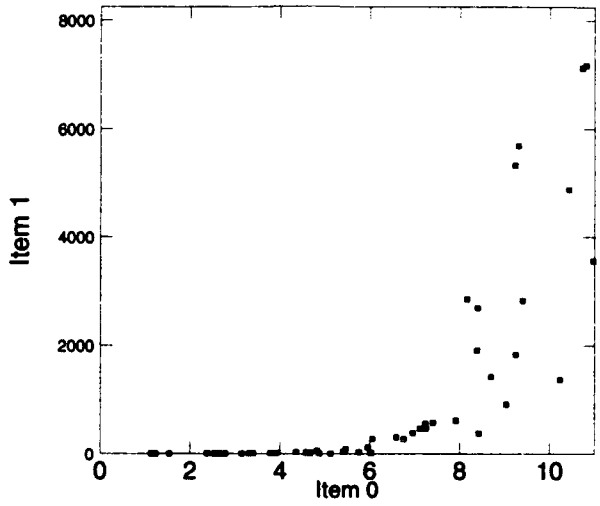


Figure 1: Exponential data.

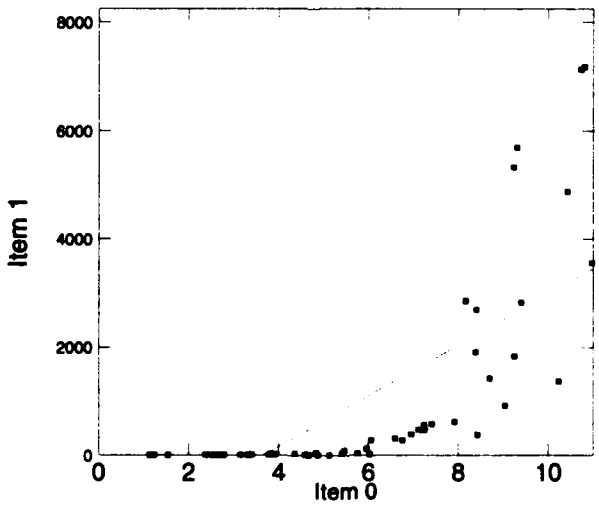


Figure 2: Exponential data with a linear fit function.

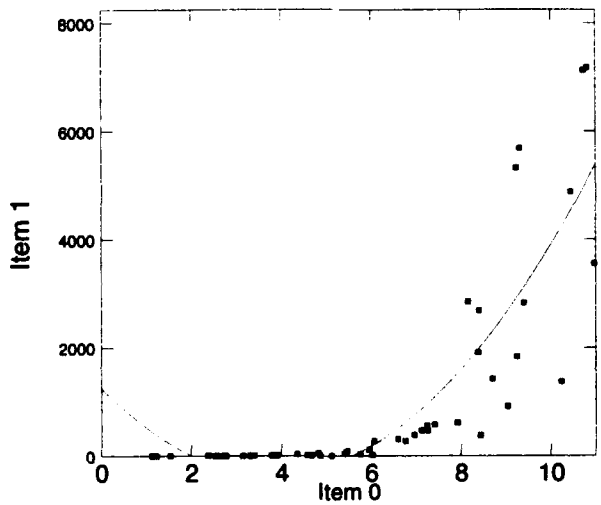


Figure 3: Exponential data with a quadratic fit function.

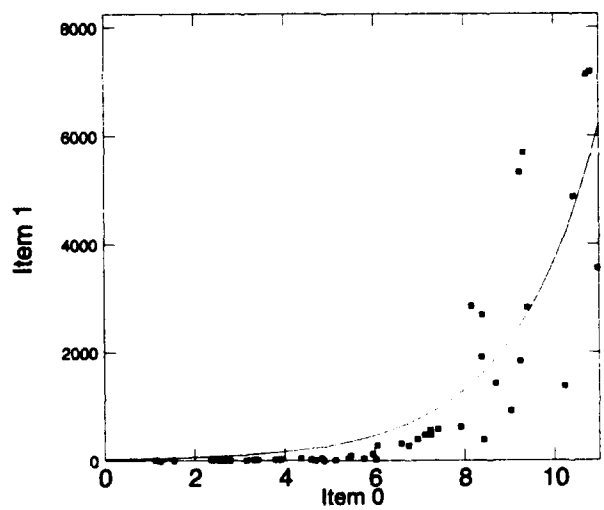


Figure 4: Exponential data with an exponential fit function.

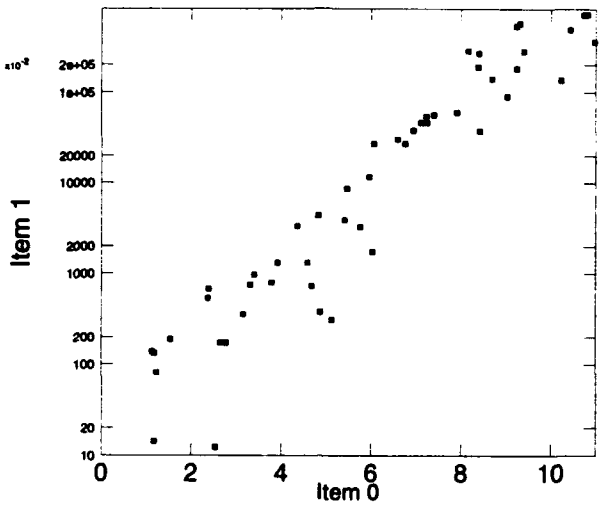


Figure 5: Exponential data plotted on a log scale.

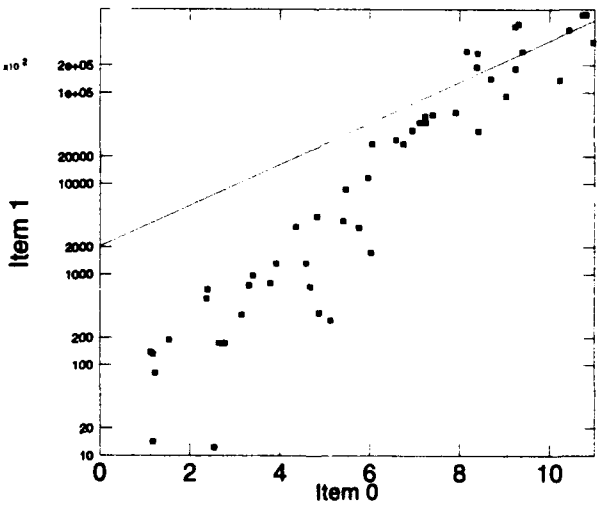


Figure 6: Exponential data plotted on a log scale with an exponential fit function.

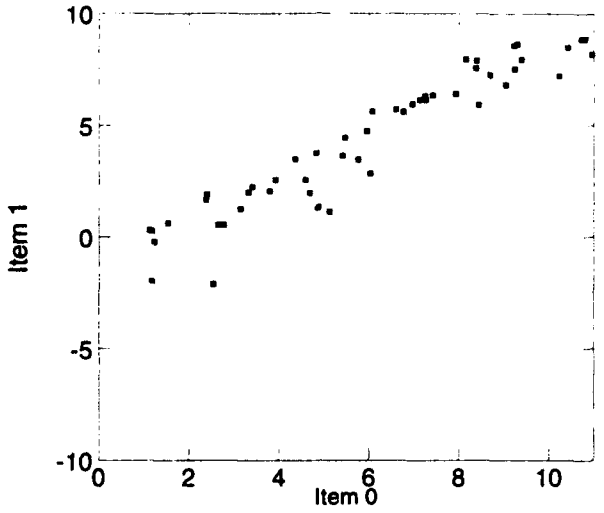


Figure 7: Logs of exponential data.

The fitness bias, toward minimizing the sum of squares, led us to a poor model. If we want to use sums of square errors, and still count percentage errors with the same weight, regardless of where they occur, we have to scale the points. This is exactly what taking logs does. (Of all data analysis techniques on earth, none is simpler or more useful than taking logs.) The logs of the data are plotted in Figure 7. Of course, it looks just like the raw data replotted on a log scale, in Figure 5, but the y-axis scale is different. This would not be significant, except that our fitting routine needs the rescaled y-axis numbers in order to do a fair job of adjudging fitness. A linear fit on log-transformed data results in the line shown in Figure 8. Comparing this with Figure 6 shows a much more satisfying model. Taking the inverse log function of our linear model will give us the best model.

Without the pictures, we might prefer the first model, the exponential fit to the untransformed data. After all, the raw data have not been tampered with, just to fit a model. (Indeed, a recent paper in Environmental Toxicology and Contamination advocated "correcting" for the process of fitting transformed data. The corrections advocated would have resulted in a fit much like our first, direct, exponential fit.) But the fitness bias of the modelling procedure has been ignored. A sum of square errors is inappropriate for exponential data, and only appropriate after logs have been taken.

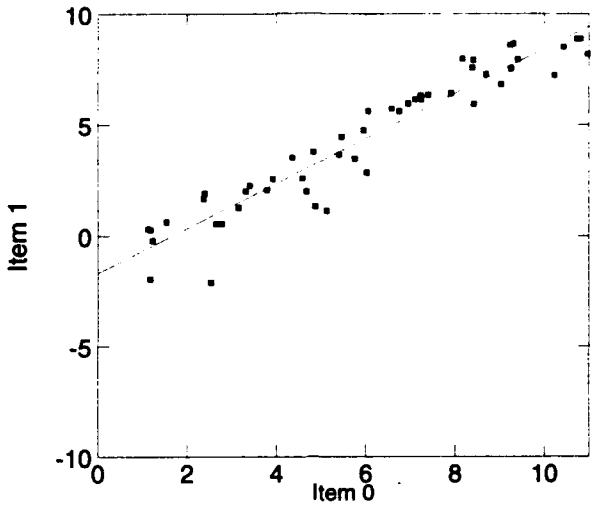


Figure 8: Logs of exponential data with a linear fit function.

3 Clustering

For the rest of this minicourse we will concentrate on the clustering model. Clustering is an extremely powerful technique, with widespread applicability in ecotoxicology. Put simply, clustering is the attempt to divide the data into several groups, or clusters. Points within each cluster should have much in common, and points from different clusters should have little in common.

Clustering hinges, then, on whether points are "similar" or "different." This raises problems in many circumstances, however. How is similarity to be defined? We have already seen that similarity should be context sensitive—in the regression example, the distance from one point to another depended on the variance of all the points. Statistical tests often differ primarily in how they measure similarity. The t-test, for instance, assumes that large differences in the mean values for the groups implies dissimilarity. The F-test, another example, assumes that small variances within the groups implies similarity within them. Each of these attempts to determine whether the within-group-similarity is significantly larger than the between-group-similarity.

Given a similarity measure, clustering still depends on the algorithm chosen. Consider the data in Figure 9, using Euclidean distance for a similarity measure. (Euclidean distance has the virtue that, in two dimensions, it is easy to gauge by eye.) Clearly, we have two groups, outlined by circles. But point A, for example, is closer to point B, in the other group, than it is to C in its own group. So, it will not do to simply say, "cluster

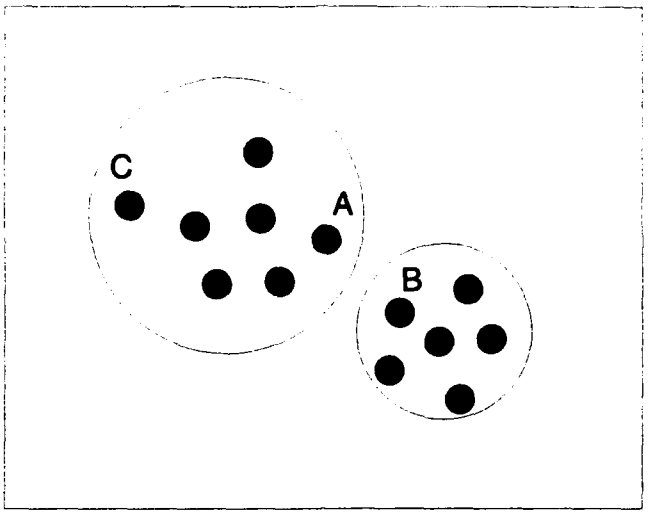


Figure 9: Two clusters of points. Yet the point labelled "A" is closer to "B," in the other cluster, than it is to "C," in its own cluster.

each point with the ones it is close to." We need an algorithm.

Later on, we will investigate a nontraditional algorithm for conceptual clustering. For now, however, consider two traditional algorithms for clustering: agglomerative and k-means.

3.1 Agglomerative clustering.

Agglomerative clustering starts with each point in its own cluster. Then, it merges the closest two clusters into a single cluster, resulting in fewer clusters, but with more points in them. For example, consider the sequence shown in Figures 10 to 14. We begin with five points which need clustering. The two closest points, point 0 and 2, are then merged to form one cluster, and relabeled with 0, as in Figure 11. In the figure, the other points are also renumbered from 0 to the total number of clusters. From this figure, the two closest points are numbered 1 and 2, so they are merged in Figure 12. Now point 2 is closer to cluster 0 than it is to cluster 1, so it is merged with cluster 0 in Figure 13, forming two clusters. If the process is continued one more step, only one cluster is left, as in Figure 14.

While this process looks simple, there are many details that must be taken into account. For instance, it is easy to tell which of two points is closer, but how do you tell which of two clusters is closer? Do you measure from the center of the cluster, or the edge? Consider the points in Figure 15, and the agglomerative process. In the beginning, the table of cluster-to-cluster distances is simply a table of point-to-point

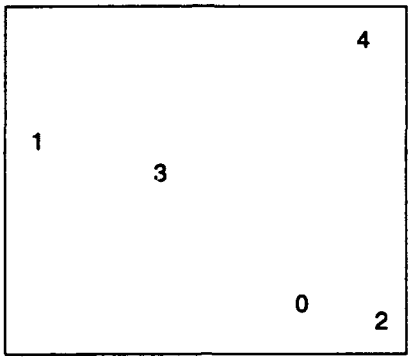


Figure 10: Agglomerative clustering, stage 1.

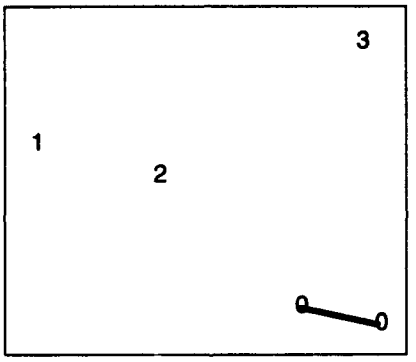


Figure 11: Agglomerative clustering, stage 2.

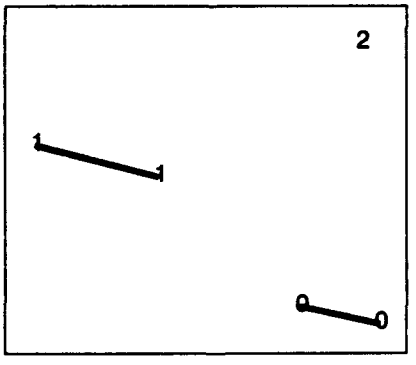


Figure 12: Agglomerative clustering, stage 3.

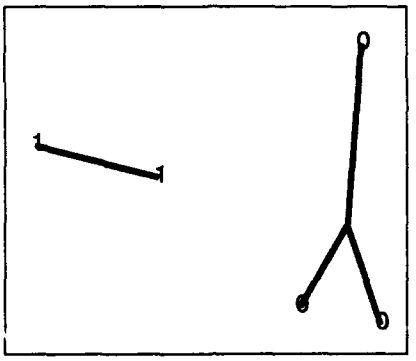


Figure 13: Agglomerative clustering, stage 4.

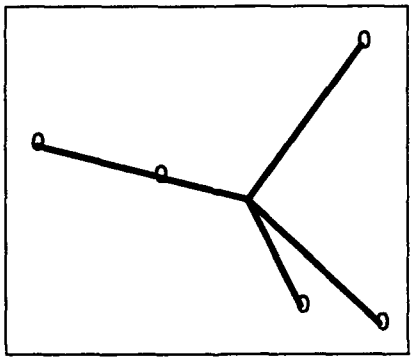


Figure 14: Agglomerative clustering, stage 5.

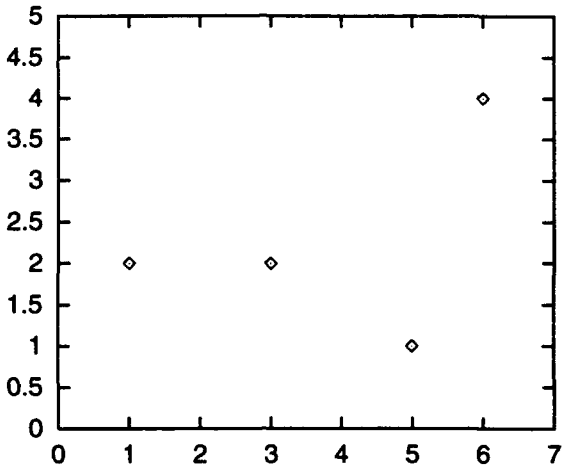


Figure 15: Points to cluster using nearest neighbor and farthest neighbor.

	1	2	3	4
1	0	2	4.12	5.38
2	2	0	2.23	3.60
3	4.12	2.23	0	3.16
4	5.38	3.60	3.16	0

Figure 16: Table of distances for points.

distances. This table is given as Figure 16. Scanning the table by eye reveals that the closest two points are 1 and 2, so we should merge them into one cluster.

Now, if we do that, how are we to fill in our table? What is the distance to *both* 1 and 2? Three possibilities suggest themselves: use the distance to 1, the farthest, use the distance to 2, the nearest, or use some average of the two. Figure 17 shows the two extreme values. Which we choose has an effect on the merging process. If we use the "nearest neighbor" strategy, then 3 should be merged with 1&2 to form the next cluster. If we use the "farthest neighbor" strategy, then 3 and 4 should be merged to form the next cluster. If we use the "mean neighbor" strategy, then 3 and 4 should be merged.

There is no clear answer to which strategy is best with agglomerative algorithms.

	1 & 2	3	4
1 & 2	0	2.23~ 4.12	3.60~5.38
3		0	3.16
4			0

Figure 17: Table of distances for points, after merging 1 and 2.

Each serves to answer the question "which points belong in a group" in a different way. Each has a different fitness bias. Nearest neighbor, for instance, tends to group together long "chains" of points. In the simple example above we saw that nearest neighbor would add a point to an existing cluster, rather than form a separate cluster. Farthest neighbor tends to avoid chains, and favors even-sized clusters. Mean or median neighbor clustering is a compromise. (The clustering illustrated in Figures 10 to 14 used means to determine nearest clusters.)

In addition to choice of distance metric and choice of algorithm, agglomerative clustering also has the problem: when do you stop? You start out with N clusters, where N is the number of points, and you end up with 1 cluster. Clearly the truth is somewhere in between, but where? The only answer to this is to have another procedure determine the "quality" of the clustering. Keep agglomerating clusters while the quality improves, and stop when it starts to degrade.

But, of course, how do you define one clustering as better than another? Some measures come to mind: a good clustering should have small within-cluster distances, relative to between-cluster distances, for instance. But how small is good? With each point in its own cluster, the within-cluster distance is zero. Should there be a penalty attached to single-point clusters? Such questions have a wealth of answers in the literature, and the answer to each must be understood before the results of an agglomerative clustering algorithm are understood. They all determine its particular fitness bias.

3.2 K-means

All agglomerative clustering algorithms are susceptible to being misled. Each makes choices about cluster membership one at a time, and, once made, the choices are never undone. There is no "big picture" in agglomerative clustering. This contrasts with K-means clustering.

K-means clustering takes an entirely different tack. Instead of clustering the points one at a time, all points are assigned to a cluster at once. To begin with, you have to choose how many clusters you want (that's the "K" in "K-means"). For concreteness, let us suppose we are looking for two. Two points are then chosen randomly as "cluster centers." In Figure 18 these points are numbered 0 and 1. All other points are now assigned to one cluster based on which "center" it is closest to. This is shown in Figure 19. Now, these points are used to find a new "center," which is the mean of all points in the cluster (the "mean" in "K-means"). Rays in Figure 19 are drawn from each point to the mean. These new centers are now used to reassign each point to the closest cluster "center." In Figure 20 you can see that some of the points have been reassigned, and the centers have moved. This process, of letting the points define the center, and then letting the center "attract" all the closest points, sooner or later settles down and no longer changes. At this point, the algorithm terminates. For the data in Figure 18, the process settled down after only three iterations. Speed of the algorithm is a notable feature of K-means clustering.

Since K-means starts with two randomly selected points, it is possible (and actually

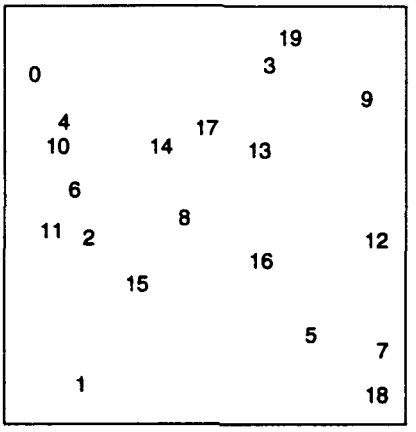


Figure 18: Kmeans clustering, stage 1.

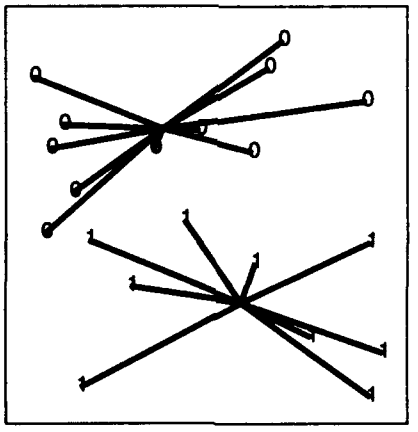


Figure 19: Kmeans clustering, stage 2.

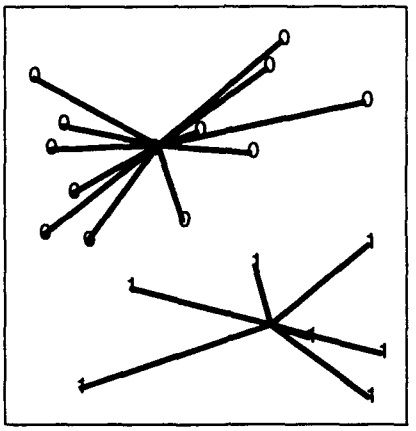


Figure 20: Kmeans clustering, stage 3.

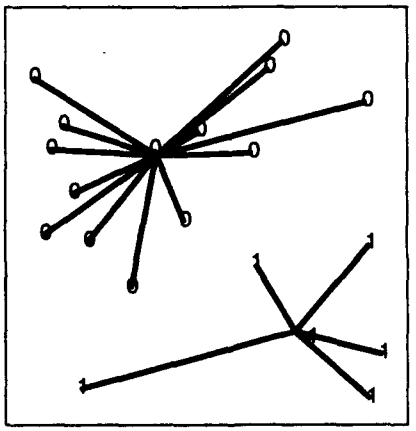


Figure 21: Kmeans clustering, stage 4.

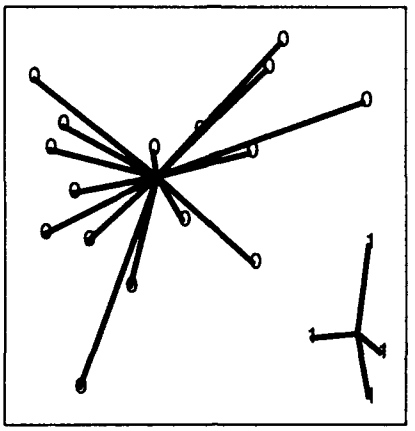


Figure 22: Agglomerative clustering of the same data from the Kmeans example.

quite likely) that restarting the algorithm will result in different clusters. The usual practice is to run the algorithm several times, and then keep the "best" clustering, where, again, some measure of clustering quality must be provided independently of the algorithm.

The same data were clustered with an agglomerative algorithm, for comparison, in Figure 22. While the overall picture is similar, there are significant differences. There are other cases where K-means and agglomerative clustering result in entirely different clusters, in spite of the fact that they are both trying to "group together points that are close."

To use the K-means algorithm, you have to know in advance how many clusters you want. If you don't know, you can always run the algorithm once for every number of clusters from 2 to 10, and then keep the best. But again, the same problem of what is "best" comes up in deciding whether two or three clusters fits best. Neither agglomerative nor K-means clustering help in this decision.

4 Conceptual Clustering

Similarity should be concept sensitive. For instance, look at Figure 23. Most people would, on first examination, say that this figure illustrates points arranged in two circles. But that would mean that the points labelled "A" and "B" belong to different clusters, even though they are the two points that, on their own, are the closest together. Even with a distance metric that corrected for variance, such as the Mahalanobis distance, these two points would still be very close (even closer, in fact).

The problem is that clustering must be both context sensitive and concept sensitive. In Figure 23, the problem is that no provision is made for the importance of "geometric figures". A specialized algorithm, using Hough transforms, for example, would have to

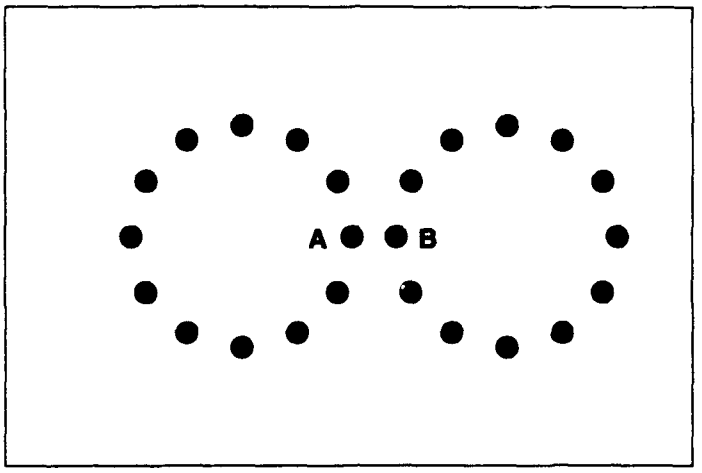


Figure 23: Conceptual clustering problem. "A" and "B" belong to different clusters, yet they are closer together than any other points.

be devised to cluster the data in the figure correctly. Geometric figures are usually not important in analyses, although there are exceptions, such as the "arch effect" noted in many ceonocline studies.

Instead of concepts based on geometric figures, or fitness based on distance measures, we adopt the following **Principle Of Nonmetric Clustering:**

TWO POINTS ARE SIMILAR IF THEY HAVE A LOT OF FEATURES IN COMMON.

How this actually gets spelled out in a computer algorithm, and the consequences for our understanding of data, will concern us for the rest of this session.

4.1 Data

Data come with many problems attached. Consider, for example, the data matrix shown in Figure 24. In this example, we assume that measurements are made on a number of species, 1...N, but the measured parameters could be chemical rather than biological. Each data point, with a unique ID 1...M, comes from a group (Grp), which could be a treatment group, or a different site in the study, and for each group we have a number of replicates (Rep).

A number of problems are evident. Some data values may be missing, noted by "???" in the table. Some may be below detection limit, noted by "BDL" in the table. Some may have a huge variance, like Species 3 (bacteria?), while some are exceedingly rare, and comprise almost all zeroes, like Species N. Also, some species may be represented by counts of individuals, while others are estimated by biomass, chlorophyll, *etc.* accounting for the inhomogeneity of numbers in the table. Finally,

ID	Grp	Rep	Species 1	Species 2	Species 3	...	Species N
1	1	1	45.7	12	???		0
2	1	2	5.3	0	2000		0
3	1	3	405.7	1.2	8000000000		0
4	2	1	???	19	???		0
...							
M	12	3	97.2	???	BDL		1

Figure 24: A hypothetical data matrix.

environmental data tables such as these are usually "broad and shallow," that is, they contain many columns, but few rows. Many species, for example, are found at each site, but budget restrictions are such that the sites can be sampled only a few times. These kinds of problems are also compounded if physical-chemical data are added to the table.

If we are to cluster data like these, the following considerations should be borne in mind:

1. The measure of similarity should not combine counts from dissimilar taxa by means of sums of squares or other simple mathematical techniques, since this will introduce a fitness bias, often with unpredictable consequences.

2. If possible, the measure of similarity should not require *transformations of the* data, such as normalizing the variance. This would also have consequences that were difficult to interpret, given species like number 3 and number N, in the table.

3. The measure of similarity should be able to work with partial data. If some samples have most, but not all, species represented with numbers, and others with "???", then we should not have to throw out the whole sample. Nor should we have to provide a pseudo-measurement for that value, such as the average of other samples. Missing data values should be overlooked. (In some cases, the fact that a data value is missing is important by itself. In these cases, "missing" should be regarded as another legitimate value, like "BDL", and included in the analysis. Usually, however, data values are missing for extraneous reasons, and can be safely overlooked.)

4. Significance of a taxon to the similarity measure should not be dependent on its size. One or several taxa with small total variance, such as rare ones, may in fact be quite significant for clustering. They may comprise a set of "indicator" species. The similarity measure should not be mislead by mere abundance.

5. Of course, any assumptions of normality, heteroscedasticity, and the usual assumptions necessary for analysis based on multivariate Gaussian distributions must NOT be made. Environmental data are almost never normal.

6. Care must be taken to avoid "over-fitting" the data. Since there are few points in a large dimensional space ($M << N$), almost any model that takes into account all species will be able to get a very close "sums of squares" fit to the few points available. A good fit in this sense should be looked at very skeptically.

4.2 The Goal of Clustering

To reiterate: **The Principle Of Nonmetric Clustering** says

TWO POINTS ARE SIMILAR IF THEY HAVE A LOT OF FEATURES IN COMMON.

While this principle sounds good, we still have a long way to go before we have an algorithm. First of all, what do we mean by a "lot" of features in common? Well, in a good clustering, under this criterion, there would be a large number, though not necessarily all, of the features (dimensions, species, *etc*) for which each cluster of points would have similar values. If height were one of these features, for example, then all the points in one cluster would be tall, another cluster medium, and a third cluster short. Of course we only mean tall (or medium or short) relative to the points in the other clusters. So what we mean by "in common" is relative to the data itself.

Further, what we mean by a "lot" of features is also relative to the data. It may be that there are no clusters with lots of features in common. Every individual data point may be completely different from all the others. In this case, clustering is clearly pointless. But if there are some clusterings where the points within each cluster have a lot of things in common, then we *prefer* these clusterings to ones where points in the clusters have less in common.

In formal terms, the clustering should create a logical description of the data. The clusters should be such that most of the points can be described by simple conjunctive descriptions involving the original parameters. For example, if a large number of the points (cluster A), in dimensions x, y, and z, had "medium", "small", and "large" values, respectively, and another large number of points (cluster B), had "large", "medium", and "medium" values on these same dimensions, then the points could be described by the two concepts:

$$\text{Cluster A:} \quad \Leftrightarrow \quad (x = \text{medium}) \wedge (y = \text{small}) \wedge (z = \text{large})$$
$$\text{Cluster B:} \quad \Leftrightarrow \quad (x = \text{large}) \wedge (y = \text{medium}) \wedge (z = \text{medium})$$

If these two sets of points comprised nearly all of the original data, then the clustering would be complete. There may be other dimensions in the original data set, other than x, y, and z, but these dimensions would be regarded as irrelevant to the above clustering if x, y, and z sufficed.

To see how this works in an algorithm, consider the series of clusterings in Figures 25 through 28. In this algorithm, points are not assigned to "nearby" clusters in any sense. Instead, points are randomly assigned to clusters, and then the quality of the

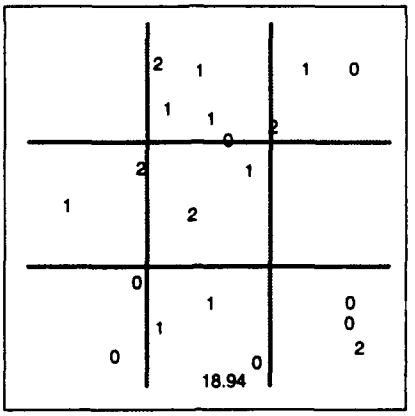


Figure 25: Riffle clustering, stage 1.

clustering as a whole is measured. In other words, do points in the same cluster have a lot of features in common? In order to assess whether they have the feature in common, each feature is divided into three regions (small, medium, and large), giving the tic-tac-toe appearance. For a larger number of clusters, more regions would be defined. If the clusters are not good, then each point, in turn, is randomly reassigned to other clusters, in an attempt to improve the quality of the clustering, until the clustering becomes as good as can be found by this process. When its as good as it can get, as in Figure 28, the process stops.

The thing to notice about Figure 28 is that in each row of the matrix, we have points from only one cluster. In other words, as far as possible, the points from a single cluster all have the same sets of properties. In the next section we will refine this notion of similarity.

Also notice that the algorithm is guided at each step by the quality of the clustering. A numeric estimate of the quality (with lower numbers meaning better) is printed at the bottom of the figures. Points are not assigned to clusters based on their nearness to other points, but on the basis of the quality of the clustering as a whole. This is why the clustering methodology is called "nonmetric clustering;" a metric is not used to determine cluster membership. Note that event the clustering methods that do use a metric still require a measure of clustering quality to know when to stop, or to determine how many clusters, or to decide between two retries.

It would also improve the clustering if we could move some points to different cells, for example, moving all the 1's to the right. Unfortunately, the position of each point on the board is fixed by the data itself, and our analysis program is not allowed to fudge the data. But, we can do something that will help. In the full algorithm, the tic-tac-toe lines are also adjusted, in the search for the best clustering. This is tantamount to adjusting the boundaries between small, medium, and large. In this way, we may improve the clustering further. We will have more to say of this, below.

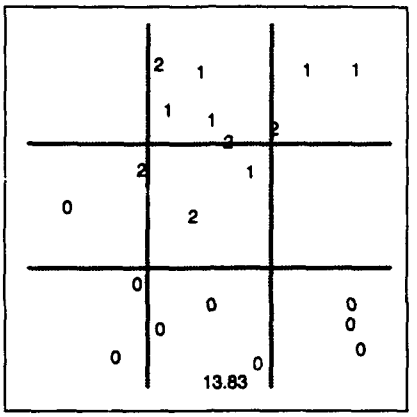


Figure 26: Riffle clustering, stage 2.

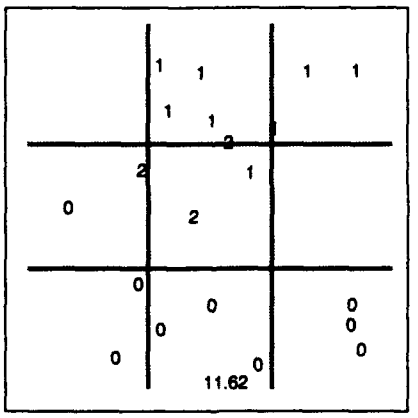


Figure 27: Riffle clustering, stage 3.

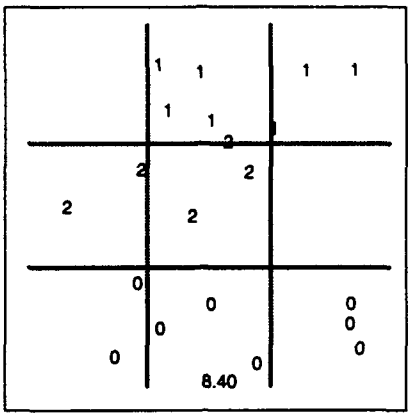


Figure 28: Riffle clustering, stage 4.

	Species 1 Counts		
	Low	Med	High
Cluster 1	3	4	10
Cluster 2	12	9	0
Cluster 3	7	11	2

Figure 29: Contingency table giving number of points (samples) with low, medium, or high numbers of Species 1 in each of three clusters.

4.3 Measures of Predictability

To formalize our intuitions, what we seek is a clustering which is informative. In a good clustering, knowing which cluster a point comes from tells us a great deal about the points features. The clustering we seek is a logical description of the data which is as informative as possible. So, we need a measure of informativeness: How much does one feature (the cluster) of an object tell us about another?

Consider the contingency table given in Figure 29. This is a pretty good clustering with respect to Species 1, since Cluster 1 indicates high counts, Cluster 2 indicates low counts, and Cluster 3 indicates medium counts. But how do we measure just how good this is? Standard measures of association in contingency tables, such as χ^2 or entropy, could be used, but we seek a more intuitive measure of the information in the table. (Entropy was used in the small demo program that generated Figures 25 through 28.)

One way to measure the information given by the clustering is to calculate the *proportional reduction in error* it gives. Suppose we had a sample, and we had to guess whether the Species 1 count would be low, medium, or high. Calculating the marginals in Figure 29 gives us 22 "low" samples, 24 "medium" samples, and 12 "high" samples. Other things being equal, we would pick the most likely value, and say that an unknown sample will probably fall in the medium range. But we would expect that prediction to

be wrong about (22 + 12) = 33 times out of (22 + 24 + 12) = 58 tries. Thus, our error rate is $33/58 \approx 57\%$, in ignorance. Suppose, on the other hand, that we are told what Cluster a sample comes from. If it is Cluster 1, then we would guess "high," Cluster 2 we would guess "low," and Cluster 3 we would guess "medium." To calculate our expected error, we need the error rate for each cluster, together with the probability for each cluster. For cluster 1, the error rate is 7 out of 17 (41%), for Cluster 2 it is 9 out of 21 (43%), and for cluster 3 it is 9 out of 20 (45%). The probability of Cluster 1 is 17 out of 58 (29%), the probability of Cluster 2 is 21 out of 58 (36%), and the probability of Cluster 3 is 20 out of 58 (34%). The expected error rate, then, is

$$(7/17)(17/58) + (9/21)(21/58) + (9/20)(20/58)$$

or about 43%. So, in ignorance we are wrong 57% of the time, while with knowledge of the clustering we are wrong about 43% of the time. So, about

$$\frac{57 - 43}{57} \approx 0.25$$

or about 25% of our erroneous predictions have been eliminated.

This quantity is the proportional reduction in error. It can be formally defined as follows, for any contingency table. Let:

- ρ_{ab} be the proportion of entries in row a, column b of the table,
- ρ_{am} be the maximal entry in row a,
- ρ_{mb} be the maximal entry in column b,
- $\rho_{a.}$ be the row marginal sum over columns,
- $\rho_{.b}$ be the column marginal sum over rows,
- $\rho_{.m} = \max_b \rho_{.b}$ be the maximal column marginal,
- $\rho_{m.} = \max_a \rho_{a.}$ be the maximal row marginal,

then the proportional reduction in error in guessing column value b is:

$$\lambda_b = \frac{\sum_a \rho_{am} - \rho_{.m}}{1 - \rho_{.m}}$$

and, symmetrically, the proportional reduction in error in guessing row value a is:

$$\lambda_a = \frac{\sum_b \rho_{mb} - \rho_{m.}}{1 - \rho_{m.}}$$

A symmetric λ can then be defined as

$$\lambda = \frac{1}{2} \frac{\sum_a \rho_{am} + \sum_b \rho_{mb} - \rho_{.m} - \rho_{m.}}{1 - \frac{1}{2}(\rho_{.m} + \rho_{m.})}$$

Some of the properties of lambda are:

1. λ is determinate except when the entire population lies in a single cell of the table.
2. Otherwise the value of λ is between 0 and 1 inclusive.
3. λ is 1 if and only if all the population is concentrated in cells no two of which are in the same row or column.
4. λ is 0 in the case of statistical independence, but the converse need not hold.
5. λ is unchanged by permutations of rows or columns.
6. λ lies between λ_a and λ_b, inclusive.

The real advantages of λ are that it is an intuitive measure of informativeness. The bigger λ is, the more one feature tells us about another. In fact, the value of λ tells us something specific: If $\lambda = 0.5$, then our errors in guessing will be cut in half. Some of our graduate students have also done some preliminary research, comparing λ to χ^2, entropy, and other measures, and have found it to be more reliable as a relative measure of association. Our nonmetric clustering algorithm, therefore, uses λ in its computation of clustering quality.

4.4 Quantitative Data into Qualitative Data

I mentioned before that the division lines between small, medium, and large are subject to adjustment. However, not all adjustments will help. Consider Figure 30. All but one of the points have been clustered into two clusters. But we have one troublesome point, labelled with an "X" in the figure. Should we put this in with the open circles, or the closed circles? Neither position helps. If we put it with the open circles, we have the following frequency table:

	X-axis		Y-axis	
	Small	Big	Small	Big
Open Circles	7	1	0	8
Closed Circles	0	4	4	0

Now the proportional reduction in error for the Y-axis is one, perfect! But the proportional reduction in error for the X-axis is

$$\frac{5/12 - ((8/12)(1/8) + (4/12)(0))}{5/12} = 0.8$$

On the other hand, if we put the "X" point in with the closed circles, we get:

	X-axis		Y-axis	
	Small	Big	Small	Big
Open Circles	7	0	0	7
Closed Circles	0	5	4	1

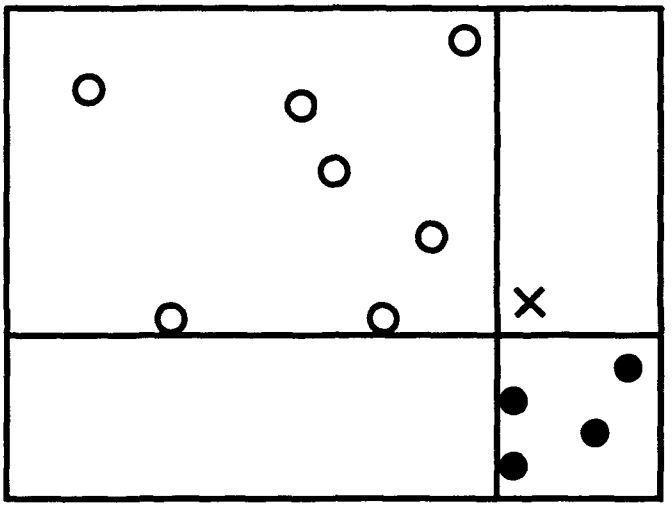


Figure 30: Quantitative data into qualitative data. The configuration of the clusters themselves determines where the division comes between "large" and "small."

Now the proportional reduction in error for the X-axis is perfect, but the proportional reduction in error for the Y-axis is

$$\frac{5/12 - ((7/12)(0) + (5/12)(1/5))}{5/12} = 0.8$$

Thus, assigning the "X" point to either the open or the closed circles results in the same λ. Under the algorithm, X will be assigned to either cluster, arbitrarily (our algorithm makes no "fuzzy" assignments).

We are not allowed to change X's position on the graph, that position comes from the data itself. However, under certain circumstances, we could fix this problem by adjusting the splits between small and large. Perhaps if we adjust the vertical or the horizontal line, we could get a better clustering. Sometimes this is possible, but in Figure 30, it is not. Moving the vertical line to the right will include the "X" with the open circles, but, alas, will create *two* troublesome points in the lower left quadrant. Likewise, moving the horizontal line up will allow the "X" to be included with the closed circles, but will again create two troublesome points in the lower left quadrant. No further readjustment of the lines between large and small can help the situation, and so the program will settle on the indicated clustering, with X assigned arbitrarily.

It is important to note, however, that the lines can be adjusted, and in the figure, already have. The vertical line has moved to the right, and the horizontal line has moved down. Clearly this led to an improved clustering, and that was why it was done. The search for a good, high-quality clustering includes the search for good definitions of "big" and "small." But the data itself constrains these definitions, too.

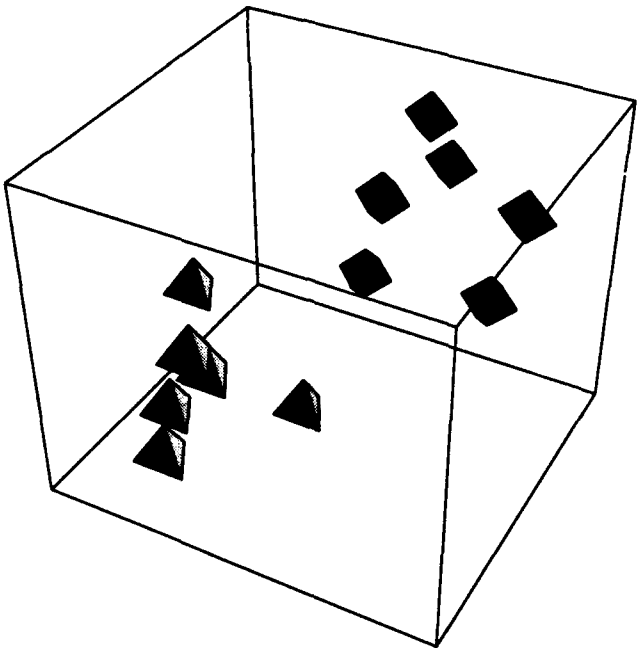


Figure 31: Control group points and treatment group points cluster when treatment has an effect.

5 Association Analysis

The clustering methodology outlined here gives the researcher deeper insight into the data. The computer program Riffle, implementing this algorithm, can tell the researcher whether or not the data form natural clusters, how strong they are, and which species (parameters) are strongly associated with those clusters.

However, given a clustering of the data (nonmetric or otherwise), it can also be used in a significance test, when treatment groups are well defined. To see this, consider Figures 31 and 32. If the treatment had an effect, we would expect points in the treated group to be different from points in the control group. In other words, we expect treatment groups to cluster well. If, on the other hand, treatment had no effect, we would not expect clustering, or, if there are clusters, they will not be associated with the treatment groups.

If we take a sample of data points, some from each treatment group, and cluster it, we will get something like Figure 33, where each point has two labels: group and cluster. In the figure we have illustrated this by shape and color. In order to tell if the treatment is significant, you only have to make a contingency table (contingency tables again!) and check for association between group and cluster. Since you are only

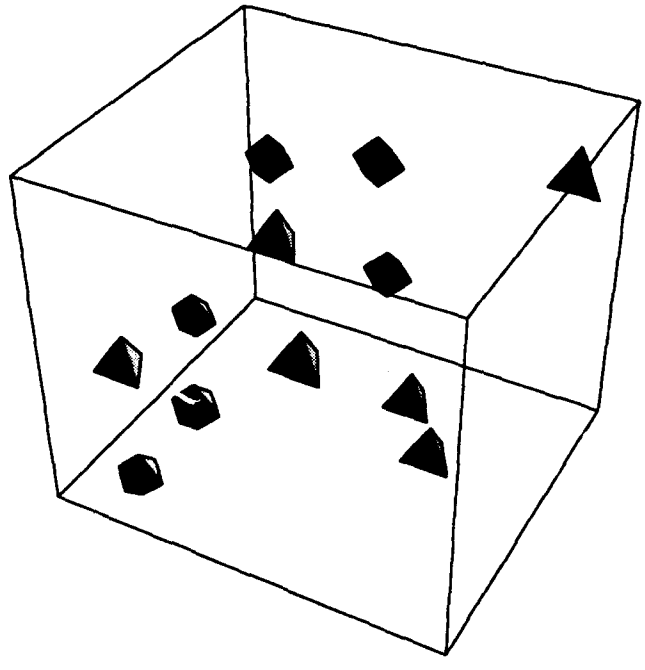


Figure 32: If treatment has no effect, we expect unclustered treatment and control groups.

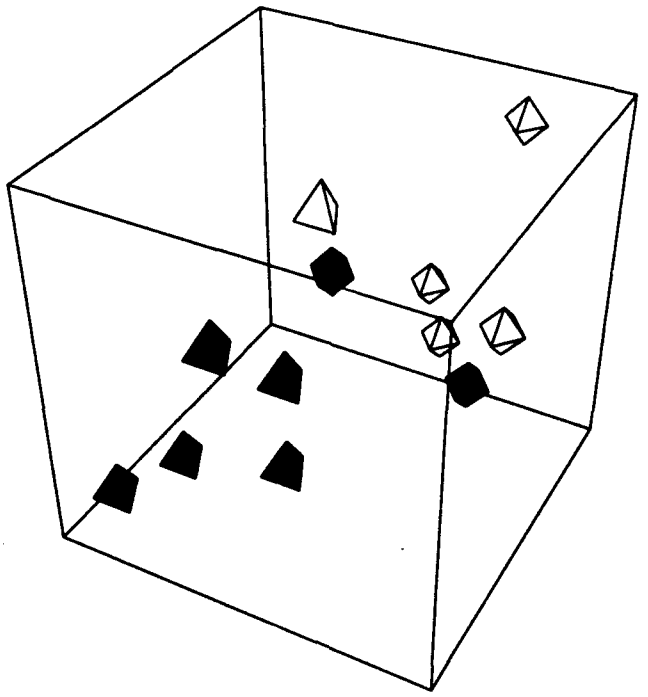


Figure 33: If points have treatment groups (shapes) and clusters (colors), the association between treatment and cluster can be measured.

looking for significance here, and not an interpretable measure, χ^2 will do fine, and the 95% confidence limit can be looked up in a table. For the tetrahedrons and octahedrons, we get the following table:

	Dark	Light
Tetrahedra	5	1
Octahedra	2	4

I'll leave it up to you to calculate the χ^2 for this example. The Riffle program calculates it automatically, if you specify treatment groups.

6 The Future

We have come a long way in our understanding of data, and why some things are similar to others. But a real challenge lies ahead. Many important ecotoxicologal studies are long-term. Environmental effects typically last for years, and undergo diurnal, seasonal, and perhaps longer (el niño?) cycles. Everything we have done so far is static: points

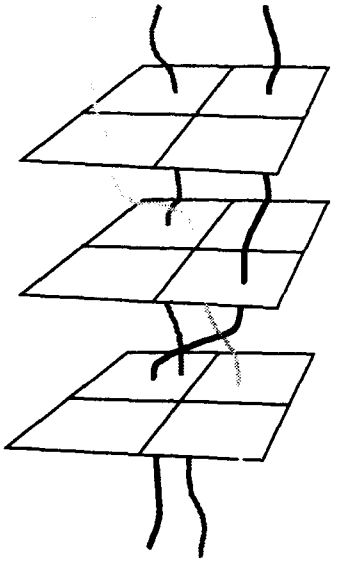


Figure 34: Spacetime worms cutting through clusterings at different times. Can temporal dynamics be understood in the same way nonmetric conceptual clustering can?

either have something in common, or they don't. But what we need is more like Figure 34. We need to understand each point as a system with a lifetime, a spacetime worm that stretches through time. Physicists have long realized the utility of this viewpoint. Modern relativity theory would be unintelligible without it, and spacetime worms and light-cones abound even in introductory texts.

If we are to make progress understanding such systems, however, we need to be more flexible than physics, with its the clockwork systems. Ecological systems can have delays, shifts, and relocations in their temporal evolution. An analysis tool that tries to compare systems over time must do more than stipulate that "temporal systems are similar if they are similar at each time." An analysis tool must be ready to do surgery to a spacetime worm, dissecting it to find its essence, and comparing this essence to what it finds in other worms. In the spirit of nonmetric clustering, nonmetric temporal clustering seeks "some times, not necessarily all times, in which the worms have some features, not necessarily all features, in common." Further, the times, for each worm, need not be the same, and the features, at different times, need not be the same.

Our analysis tool to do this kind of surgical search over spacetime worms is still in its infancy. We call the program Riggle, but that is a topic for another seminar.

Appendix 1
Software

All software described in this manual and used in the presentation is available by anonymous ftp from iceberg.cs.wwu.edu in the directory ~ftp/pub/matthews.

Most of this software has a graphical interface which requires the NEXTSTEP operating system, available for 486 PC's from NeXT, Inc. However, the source code for a simple, command-line version of Riffle is also available at the above location.

Included with this manual is a DOS diskette containing the source code for the command-line Riffle program. NOTE: this code failed to compile correctly under Turbo C++ and Turbo C++ for windows. The author does not believe this program is reliable under DOS. Also on this disk are two recommended statistical and graphical packages for MS windows, xlispstat and gnuplot. These are copyrighted programs by other authors, but are freely distributable. All software is provided "as is," without warranty and is used at your own risk.

```
*****************************************************************************
* Riffle
*
* Author: Geoffrey Matthews
*
* Permission to use, copy, modify, distribute, and sell this software and its
* documentation for any purpose is hereby granted without fee, provided that
* this permission notice appear in supporting
* documentation, and that the name of Geoffrey Matthews not be
* used in advertising or publicity pertaining to distribution of the software
* without specific, written prior permission.  Geoffrey Matthews
* makes no representations about the suitability of this software for any
* purpose. It is provided "as is" without express or implied warranty.
*
* GEOFFREY MATTHEWS DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS
* SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS,
* IN NO EVENT SHALL GEOFFREY MATTHEWS BE LIABLE FOR ANY SPECIAL,
* INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM
* LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE
* OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR
* PERFORMANCE OF THIS SOFTWARE.
*
* RIFFLE AUTHOR:
*   Geoffrey Matthews
*   Computer Science Department
*   Western Washington University
*   Bellingham, WA 98225
*   (206) 650-3797
*   matthews@cs.wwu.edu
*
*****************************************************************************
```

Appendix 2

Riffle User's Manual

Technical Report: Development of a Nonmetric Clustering User Interface

Michael J. Roze
Computer Science Department
Western Washington University

November 11, 1993

Contents

List of Figures

1 Introduction

This paper describes a user-friendly interface to the RIFFLE nonmetric clustering program [5]. Following a brief introduction to clustering, and to the context of the research project, this paper will describe how to use the interface and in the process will explore some of its capabilities.

Clustering is a data analysis technique that attempts to fit the data to a number of clusters, or subpopulations, each with distinct properties. Clustering algorithms attempt to group the data points by maximizing within cluster similarity and simultaneously minimizing between cluster similarity. RIFFLE implements a new approach to *nonmetric clustering* developed by Matthews and Hearne [5].

The object of clustering is to identify sub-populations in the data. The clusters found may verify a researcher's ideas about the data set, or may help the researcher to formulate new hypotheses about the data. The strength of clustering algorithms is their ability to sift through data sets that are too large for human consumption, and (depending on how the programs are written) can avoid personal biases that can interfere with research done by humans.

It is appropriate to make a distinction regarding the terminology used in this paper.

- *Group* is used to mean sub-populations in the data that the researcher knows about in advance of the clustering analysis. These are independent variable such as, the dosage of toxicant administered, gender, or season.

- *Cluster* is used to mean the sub-populations in the data that the clustering algorithm finds. The clusters may or may not be similar to the groups, and may or may not tell the researcher something new about the data. An association analysis is used to check for similarity between groups and clusters.

In RIFFLE'S original, command line, form the program requires the user to have a fairly thorough understanding of the input arguments and their effects. However, one of the goals of the research project is to make this statistical research tool accessible to researchers in a wide variety of disciplines. This motivated a project to develop a graphical user interface to the RIFFLE clustering program, making it easier to use, and giving graphical results in addition to the original text output.

The interface is implemented using the NeXTSTEP operating system and software development environment. This platform can be run on NeXT machines and computers with Intel '486 processors.

2 Using the Interface

2.1 Data Files

A brief discussion on data sets is included below. See the document *Formatting Guidelines for Riffle Data Sets* for a thorough discussion on data sets and description files.

Input files are expected to be real numbers or integers separated by white space (spaces or tabs). The data file should have a format similar to Figure 1.

RIFFLE is able to accept data with missing values when they are explicitly indicated by values less than or equal to **-99**. In contrast, the regular data values must be **zero**, **positive integer**, or **positive real** numbers, but not negative numbers. This means that data sets with values less than

```
1   5.1   3.5   1.4   0.2
1   4.9   3.0   1.7   0.2
1   4.7   3.2   1.3   0.4
⋮    ⋮     ⋮     ⋮     ⋮
2   4.5   3.8   1.0   0.3
⋮    ⋮     ⋮     ⋮     ⋮
3   5.2   3.0   1.8   0.5
⋮    ⋮     ⋮     ⋮     ⋮
```

Figure 1: Data file with three groups.

zero must be adjusted so that the smallest value is greater than or equal to zero. Data values may appear in scientific notation format. If a data point is missing the x or y feature (or both) it will not be plotted. While RIFFLE can accommodate data that has some missing values, it probably does not make sense to run the interface with data that, for example, is missing half of its points.

In addition, the association analysis (Section 2.6), and plot by group (Section 2.7.4) features require the data file to either have an attribute listing each point's group, or to be structured in a way that allows the interface to distinguish the groups,

- all of the points in a group must be listed consecutively in the data file, and
- all groups must have the same number of points.

With three groups, the interface will consider the first third of the data points as group "1", the next third as group "2", and the last third group "3" (as in Figure 10).

When the interface is grouping the points this way and it notices that the number of groups does not evenly divide the number of data points, it reports this situation. Once the warning is acknowledged the interface will continue as best it can. If the number of points in each group is unequal, and the groups are not labeled, then using the plot by group feature is not recommended as the interface may give unreliable results.

The input data file can be selected by choosing the menu items "File" then "Open DATA" (Figure 2). This opens the file viewer window so the user can select an input file. The interface opens the file viewer automatically if the user forgets to open the input file before starting computations.

2.2 Description Files

A brief discussion on description files is included below. See the document *Formatting Guidelines for Riffle Data Sets* for a thorough discussion on data sets and description files.

Optionally, a description file can be provided. This file lists the individual features, naming them, directing the application to treat them as discrete or continuous, and directing the application to include or exclude them from the analysis.

Each feature name is required to be a string without blanks. As an example "Sepal_Length" with an underscore separating the words is acceptable, but "Sepal Length" separated by a space

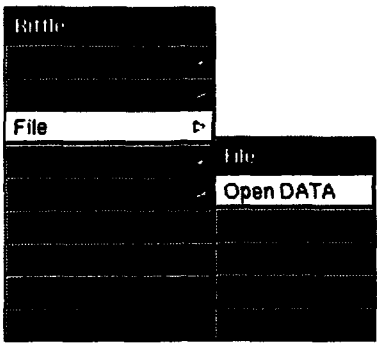


Figure 2: Main menu and File submenu.

is not acceptable. The file should consist only of feature names, one of the words "**exclude**" or "**include**", and one of the words "**continuous**" or "**discrete**" on each line.

The file may optionally include the word "**grouptag**" once to identify the attribute holding the group information. The grouptag attribute is automatically excluded from the clustering analysis so that RIFFLE is blind to the groups. The file may also include the string "**numberofgroups: n**" where **n** indicates the number of groups for the association analysis independent of the number of clusters.

Omitted descriptions default to "**include**" and "**continuous**". If "**numberofgroups: n**" is omitted, the number of groups defaults first to the number of groups observed in the **grouptag** attribute if it exists. Otherwise, number of groups defaults to the number of clusters sought.

Likewise, when the **grouptag** is available that feature will determine which group each point belongs to. When **grouptag** is not available, the points are expected to be listed in the data file such that group "1" consists of the first $\frac{1}{n}$ data points listed in the file, group "2" the second $\frac{1}{n}$ data points, and so on where n is the number of groups as determined above. Here is an example description file which corresponds to the data file in Figure 1.

```
numberofgroups: 3
Group_Number                exclude   grouptag
Sepal_Length    continuous  include
Sepal_Width     continuous  include
Petal_Length    continuous  include
Petal_Width     continuous  include
```

Figure 3: Description file with five attributes, four *included* attributes, and three groups identified by the first attribute "**Group_Number**".

If no description file exists, the number of features is set equal to the number of data values in the first line of the data file, in which case all features are taken to be continuous and all are included.

The description file is opened by the file submenu item "Open Desc". Description files can also

be closed or saved by items on the file submenu. Closing a description file will scan the data file to determine the number of features, reset the feature names to the defaults "`Attr1`", "`Attr2`", and so on, and will reset the features to be included and continuous. This is appropriate when using a new data file that is not accurately represented by the current description file. Saving the description information to a file allows the user to retrieve that information when using the same data set or one with identical format.

2.3 Input Arguments

When performing a clustering computation the program requires information from the user. The Arguments window (Figure 4) accepts the user's choices, starts computations, graphs the data and results, and reports on the computation's statistical significance.

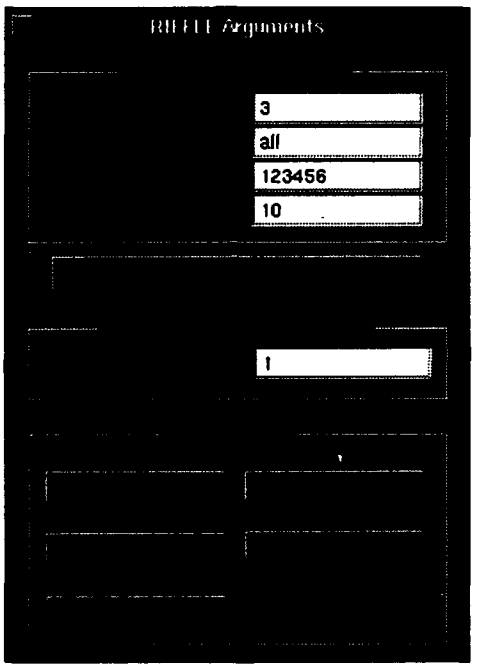


Figure 4: Arguments window.

2.3.1 Number of Clusters

The first argument is the number of clusters the program will fit the data to. A researcher may, or may not, know how many clusters are appropriate. Performing the computations with different numbers of clusters can give the user a feel for whether the data can be usefully described by clusters, and if so, how many clusters. The interface's text output (Section 2.5) shows the average quality of each clustering run. This gives the researcher an idea of how many clusters are most appropriate based on which number of clusters has the highest average quality value.

The researcher must, however, analyse the data further to verify the initial findings. For example, suppose a researcher starts with a value of three, then proceeds to four, five, and six clusters. The interface may show four clusters as having the highest quality, suggesting that there are four sub-populations in the data. However looking at the data may show that there are obviously two clusters. How can this happen? Since four is the closest multiple of two, there is a good chance that four clusters will also show a strong quality, and may be misleading.

2.3.2 Significant Features

Significant features tells the program how many of the features to consider in the computation. The argument "all" instructs the program to consider every included data feature in the average quality measure. By giving a value less than the number of included features the program can choose the features it thinks are the most important. For example, if there are six features, the researcher can ask the program to choose the best four features (Significant Features: [4]). This has the effect of excluding from the computation the two features that the algorithm finds contribute the least to the proportional reduction in error (PRE)—the quality measure RIFFLE uses for clustering [5].

The significant features argument exploits the program's ability to decide which features are the most noisy (or random) in relation to choosing sub-populations in the data. These noisy features can then be excluded from consideration allowing the program to focus on features that seem to do a better job discriminating clusters. This type of analysis can give researchers a new perspective on which attributes may be the most important.

2.3.3 Random Seed

Each time a computation is run with the same information in the arguments window, and the same input file, the results will be identical. By changing the random seed the user can force the program to use a new set of pseudo-random numbers causing the results of the next computation to be different. (See Section 2.3.4).

2.3.4 Number of Retries

The clustering job that RIFFLE is attempting to perform is an enormous task. Systematically checking every permutation of data points across the clusters would result in a computation that takes an intolerable length of time for all but the smallest data sets. In light of this, clustering algorithms, including RIFFLE, make approximations to this ideal. RIFFLE uses pseudo-random numbers to place the points in initial clusters, then proceeds to rearrange the points until a local best clustering case is arrived at. The "Number of Retries" is the number of times RIFFLE is to perform this analysis, each time keeping the results only if the overall quality was better than the previous best.

So what values are appropriate inputs? Ten retries usually gives excellent results, and good results can be obtained in five or fewer retries for a quick analysis. If, for instance, you wish to simply look at the data plots with less emphasis on how well the points are clustered, then one retry is all you need. Noisy data may have clustering results (average clustering quality) that continue to improve with retry values above ten, this can be investigated on the data sets in question.

2.4 Computing

The Compute button will begin the computation. If the input file has already been selected the computation will start right away. If the input has not been opened, the interface will open a file viewer window that allows the user to select the input data file. When the computation is done the output will be displayed in text form in the Results window (Section 2.5), the association analysis values will be displayed (Section 2.6), the Features window will be updated (Section 2.8), and the results will be plotted (Section 2.7).

2.5 Text Results

The clustering results are displayed in the Results window as shown in Figure 5. In this example the data file "iris.dat" was analyzed. The next line indicates that 150 data points described by four included features were placed into three clusters, using all four features in the clustering analysis. The text also show the number of retries and the random seed used.

The text results also report the number of features that the program found suspicious of having degenerate data (not shown in Figure 5). A degenerate feature has an excessive percentage of either identical data values, or missing data values. When these instances are found, the program does not use these features in the computation, but marks them as excluded in the features window and flags them as degenerate in the text results.

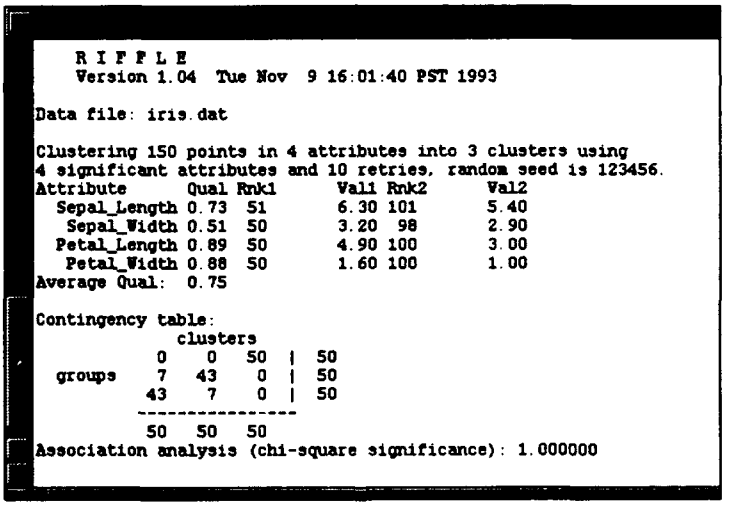


Figure 5: Results window showing text output.

Qual is the quality of the feature in the current clustering analysis, where quality is the proportional reduction in error (PRE) value discussed in Matthews and Hearne [5]. The quality value ranges from 0.0 to 1.0, with 1.0 being the highest quality. If the program is run with fewer significant features than included features then a value of 0.0 will be shown in the quality column of the non-significant features.

The ranks (**Rnk**) are the split point positions in a list of the values of that feature sorted in

descending order. The values (`Val`) are the actual numeric values of the data point at that rank. For example, in Figure 5 the Sepal_Length line indicates that the first split for that feature is at the data point with the 51st largest sepal length (out of the 150 member sample), and the actual data value for that split point is `6.30`. Likewise the same line shows the second split is at the `101st` largest sepal length, and the actual value is `5.40`.

The average quality value (`Average Qual`) is the average of the above quality values. This value gives some indication of the strength of the clusters found. With fewer significant features, the average quality value has a better chance of being close to 1.0. Our experience is that average quality values above 0.50 (with eight or more significant features) frequently indicates that RIFFLE has found some convincing sub-populations in the data. This does not, however, guarantee that the clusters match the groups. Perhaps the clusters have nothing at all to do with the groups. It is also true that with fewer included features, and fewer significant features, the average quality value will tend to be higher. Fast rules cannot be given for this type of analysis. There is no substitute for having an expert in the appropriate discipline experiment with specific data sets and arguments.

If we are looking for two clusters then one split point is defined, and the results show one rank column and one value column. If we are looking for three clusters then two split points are defined, each split point having one set of rank and value columns.

Next, the contingency table and the corresponding χ^2 (chi-square) statistic for group/cluster association are listed (discussed in Section 2.6).

If some features are excluded by the description file, or by the features window, those features will not show up on the text output at all.

Menu items in the Interface are available to print the text results to paper or save them to a file.

2.6 Association Analysis Results

The association analysis statistic appears in the arguments window (Figure 6), and in the text results window with its corresponding contingency table (Figure 7).

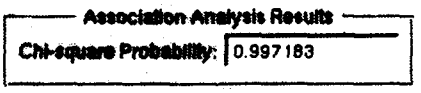


Figure 6: Association analysis χ^2 statistic (from arguments window).

The interface constructs a contingency table with the *known groups* on the y-axis and the *computed clusters* on the x-axis. Then it computes the χ^2 (chi-square) statistic to measure the significance of the association between the groups and the clusters [9]. The null hypothesis is that groups and clusters have no association. In this case the probability of a particular value of cluster number given a particular value of group should be the same as the probability of that value of cluster number regardless of group. χ^2 tells us at what significance level the null hypothesis is rejected (or more correctly `1 - significance level`). Large values of probability indicate a significant association. Values that are above 0.99, for example, indicate that the association is

```
Contingency table:
                clusters
            4   1   0   1  |  6
            0   5   1   0  |  6
  groups    3   1   1   1  |  6
            0   0   5   1  |  6
           --------------------
            7   7   7   3
Association analysis (chi-square significance): 0.997183
```

Figure 7: Contingency table (from text results window).

significant at the 99% level, i.e. that there is a less than 1% probability that this level of association would happen by chance.

In order for the association analysis to give meaningful results the interface must have the information linking each point to its group. Section 2.1 discusses the requirements for formatting the data sets so that this information is available.

Note that the probability values can be shown in scientific notation (i.e. 1.391489e–42) and are always between zero and one.

2.7 Graph Results

The interface also displays the clustering results graphically. This is a way of representing, at the same time, both the input data and the clustering results.

The interface can graphically display the data by either a scatterplot, or by a scatterplot matrix, and by cluster, or by group.

At this time, all of the graphs have the property that more than one point can be at the same location, causing the symbols to be plotted one on top of the other. This may result in the lower point being obstructed from view, and may result in plots that have fewer visible data points than expected. This may also result in unusual symbols. If, however, the symbols are exactly the same, or even just the same shape, then only the last symbol plotted will be visible. The same is true for plots that use numerals and letters to represent points. If different numerals are plotted at the same location, they will result in an unusual symbol. But if two occurrences of the same numeral occupy the same location they will be indistinguishable from a single point.

In order for the data points to be plotted they must be "included" in the Features window (see Section 2.8).

As with the text results, menu items in the interface can print the plots to paper, or save the plots to an encapsulated PostScript file (.eps).

2.7.1 Scatterplot

The plot option shows any two included features graphed against each other in two dimensions (Figure 8). The x and y features are selected in the features window (Section 2.8). At the end of the computation the interface selects the two best features (those with the greatest PRE values) and sets the best feature to the x-axis and the second best to the y-axis.

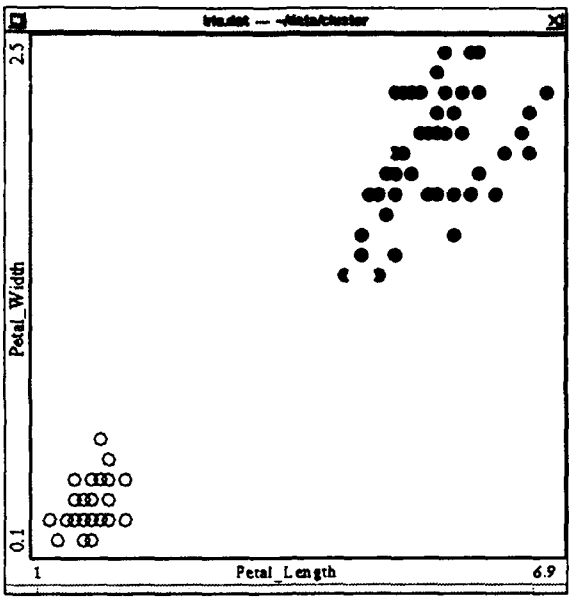


Figure 8: Scatterplot of Anderson's iris data showing three clusters.

Plotting the data and cluster results can help the researcher discover whether, and how well, the data fit the clusters. Figure 8 suggests a fairly strong clustering since both x and y features predict almost perfectly which group a point is in.

The vertical and horizontal lines show the split points for the clustering. Figure 8 shows two split points on each axis because three clusters were requested. At least one point will always fall on each split. When a point lies on the split it indicates that the point belongs to the region above (if the line is horizontal), or to the right (if the line is vertical). In Figure 8 the symbols on the bottom and left edges of the center cell are shown on the split lines, but are in fact included in the center cell of the grid.

It should be noted that there is a limit to the number of symbols the interface can plot.

2.7.2 Scatterplot Matrix

The scatterplot matrix shows several features plotted against each other in two dimensions (Figure 9). At the current time up to six features can be included in the matrix. As it does for the scatterplot, the interface finds the best set of features, based on the PRE values, and displays them in the scatterplot matrix. Other features can be selected with the features window and the matrix will automatically resize to accommodate fewer or more features (up to the maximum), without changing the window size.

The S.W. to N.E. diagonal is filled with text cells. Each text cell indicates that plots on the same **row** use that feature on the **y-axis**, and plots on the same **column** use that feature on the **x-axis**. Centered vertically and horizontally in the text cell are the feature name and quality. In the S.W. corner of each text cell is the input file's minimum value for that feature, and in the N.E.

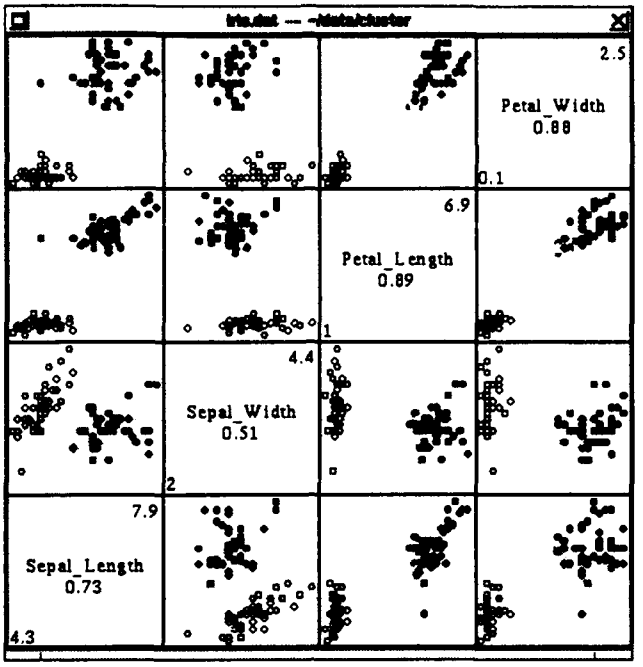


Figure 9: Scatterplot matrix showing all four Iris features.

corner is the maximum value.

The example in Figure 9 shows that the two features, petal length and petal width, are predictive, with a quality value close to the maximum of 1.0. It also shows that the sepal width feature does not contribute much to this particular clustering since the data points do not separate into discernible groups along that axis in the matrix. This visual weakness reaffirms the feature's low quality value.

Observe that the scatterplot matrix's top row, third column is the same plot as that shown in Figure 8, except that it is scaled differently (fitting the matrix into the given window dimensions).

2.7.3 Plot by Cluster

Plotting by Cluster is demonstrated in the two sections above. The plots show how RIFFLE places the data points into clusters. Each data point is plotted at coordinates equal to two of its features. The point's cluster determines the symbol used to represent it. This gives the researcher an idea of how well the clusters represent spatial sub-populations in the data.

Plotting by cluster represents each cluster by a different geometric symbol or color, whereas plotting by group represents each group by numeral.

2.7.4 Plot by Group

Figure 10 shows the same data as the plot of Figure 8, except that the plot by group option is used instead of the plot by cluster option. The points in plot by group are indicated by numerals instead

of the geometric symbols used when plotting by cluster. Other than the labels, the graphing is done the same way in both plots so a direct comparison is appropriate. In fact, it is anticipated that researchers will swap between the two plot types to check for differences between the groups and clusters. When the color display is used, plotting by group will show the groups as numerals, and the clusters as colors. This way, a direct comparison can be made between the groups and clusters without redisplaying the plot.

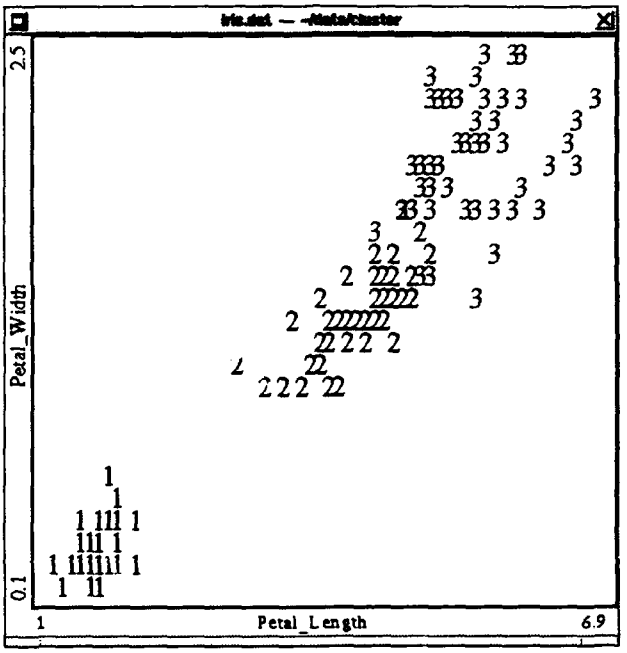


Figure 10: Plot by group

Graphing the data and results with this option is a tool to help answer the question, "Now that I have the results of the clustering, how closely do they match the subpopulations that I know exist?" In the Iris example there would be three groups which correspond to the three types of irises studied. The question is, "Do the clusters do a good job of grouping the data by the type of iris?" Comparing the group and cluster plots can help to answer this question.

It should be pointed out that the RIFFLE program is "blind" to the groups. That is, RIFFLE assigns points to clusters without any knowledge of which group the points come from. However, the data is plotted by the interface, not RIFFLE. The interface uses a group label attribute or structure in the input file to show the groups while RIFFLE remains naïve to that information.

In order to plot by group, the interface must be able to distinguish between the data groups as discussed in Section 2.1.

2.7.5 Adjustable Symbol Size

Symbol size in the plots can be adjusted with the menu item "Format" and submenu "Symbol Size". The best size will typically depend on the number of points in the dataset, and the size of

the plots (which can be changed by resizing the window). Using one size for the scatterplot and a slightly smaller size for the scatterplot matrix seems to work well.

2.7.6 Printing Graphs

The Print menu item has both "Graph", and "Full Page Graph" options. The Graph option will print the graph window at it's current size. If the window is larger than one page then the printing process gives unpredictable results. Otherwise the printed graph will be approximately the size seen on the screen. The Full Page Graph option resizes the viewing window to page size and directs the output to the printer. This option will make printed graphs with the graph scaled to page size, and would be useful for making printed graphs that are always the same dimensions (i.e. not dependant on how you resized the graph window in that particular session). It is best to choose portrait or landscape page orientation with the Format menu item prior to using the Full Page Graph option.

2.7.7 Color

The color option allows the graphs to display different clusters by color, always using the same symbol shape. Even if color is used on the screen, when printing, the interface will adjust the symbols to accommodate a non-color printer.

2.8 Features Window

The Features window provides on the fly choices paralleling those made with the description file, and also controls which features are graphed in the scatterplot and scatterplot matrix.

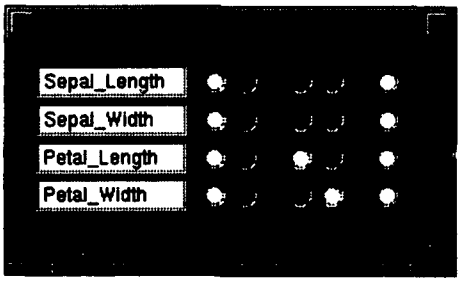


Figure 11: Features window: highlighted buttons (white) indicate features included in computation, and features to graph in plot (x and y) and plot matrix (matrix).

The features window allows the user to change the feature name, include or exclude the feature in the computation, and designate the feature as either continuous or discrete. The features window also controls which features are plotted in the scatterplot and scatterplot matrix graphs. Columns "x" and "y" allow exactly one of the features to be selected at any time. These columns direct which features are plotted on the scatterplot's x, and y axes. The "Matrix" column, on the other hand, allows two or more features to be selected, and will plot these in the scatterplot matrix (up to the maximum).

Figure 11 shows all four features included in the computation, all features are continuous, the third and fourth features are selected for the scatterplot (columns x and y), and all features are selected for the scatterplot matrix (column Matrix).

A feature must be **included** in the computation in order for the interface to plot it. The interface will edit for this requirement and deselect the feature if it does not qualify for plotting. This will cause an error panel to appear, and the graph will be cleared.

The information in the Features window (feature name, whether it is included or excluded, and whether it is continuous or discrete) can be saved to a description file (Section 2.2) by the menu "File", submenu "Save Desc".

3 Future Plans

- Although clustering has historically been considered an exploratory data analysis technique, the research team is investigating promising applications of the nonmetric clustering tool for predictive statistics as well.
- The team is developing an interface version that includes tools for performing a broader cross-analysis of treatment group type data with several statistical techniques including the RIFFLE algorithm.
- The research team is also experimenting with running RIFFLE on diverse classes of data sets to see what insights this clustering technique can give into new and classic data analysis problems.

4 More Information

For more information about nonmetric clustering, the RIFFLE program, or their applications refer to these papers [5, 6, 7, 8, 10]. Questions about the interface, or the above issues can also be directed to:

Geoffrey B. Matthews: `matthews@skipjack.cs.wwu.edu`
Michael J. Roze: `roze@rum.cs.wwu.edu`.

5 Acknowledgement

This research is supported by USAFOSR Grant No. AFOSR-91-2091 DEF.

References

[1] E. Anderson. The irises of the Gaspe Peninsula. *Bull. Amer. Iris Soc.*, 59:2–5, 1935.

[2] E. Anderson. The species problem in Iris. *Ann. Mo. bot. Gdn.*, 3:511–25, 1936.

[3] E. Anderson. A semigraphical method for the analysis of complex problems. *Proc. of the Nat. Academy of Sciences*, 13:923–27, 1957.

[4] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Ann. Eugen.*, 7:179–88, 1936. Reprinted in Fisher, R. A. (1950). em Contributions to Mathematical Statistics. Wiley, New York.

[5] Geoffrey Matthews and James Hearne. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):175–184, 1991.

[6] Geoffrey Matthews and Robin Matthews. A model for describing community change. In *Pesticides in Natural Systems: How Can Their Effects Be Monitored?*, Environmental Research Laboratory/ORD, Corvallis Oregon, 1990.

[7] Geoffrey B. Matthews, Robin A. Matthews, and Barnard Hachmöller. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Canadian Journal of Fisheries and Aquatic Sciences*, forthcoming, 1991.

[8] Robin A. Matthews, Geoffrey B. Matthews, and William J. Ehinger. Classification and ordination of limnological data: a comparison of analytical tools. *Ecological Modelling*, 53, 1991.

[9] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1988.

[10] Eric P. Smith, Kurt W. Pontasch, and John Cairns, Jr. Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.*, 24(4):507–514, 1990.

Appendix 3

Formatting Guidelines for Dataset Preparation

Formatting Guidelines for RIFFLE Data Sets

Michael J. Roze

November 11, 1993

This document describes issues in designing and formatting a data files for analysis with the RIFFLE non-metric clustering program. These guidelines also touch on some issues in designing experiments which lend themselves to statistical analysis.

Here's a brief clarification of the terminology used below:

- The term "Point" is used to describe a specific subject in the experiment (i.e. a microcosm existing in a particular flask, a mouse, or a geographic region). In terms of a data file, each point is a row in the file, and each row consists of a specific number of numeric values.
- "Attribute" or "feature" is some measure that describes a point at a given time (i.e. number of small daphnia, temperature, or percent ground cover). Several attributes make up each row, and in terms of the whole data file the attributes can be considered columns.
- "Data set" or "file" describes an ASCII character file that contains information on a group of points (typically at a specific time). Each of the points (rows) in this data set must have an identical number of attributes describing it.

Data Set Organization

The data file must be a simple ASCII file without special characters or formatting. The file should have only numeric data and whitespace.

If the file is taken from a spreadsheet some manipulation is typically required to simplify the data set. The column headings need to be removed from the file, as do comments and extraneous numeric values (such as the date). Spreadsheets on some computer systems also place special characters such as `^M` or `^Z` in the data files even when they are saved as text or ASCII. These special characters (which are invisible in some text editors) must be removed prior to running RIFFLE.

The best strategy for constructing useful data sets is to record the data sheet information in a simple ASCII text file (such as Figure 1), and then have a spreadsheet read in that file. This will encourage the use of simple data files (which can be processed by RIFFLE, SAS and other statistical programs), while the spread sheet itself can store the column headings, dates, and comments that make the data more user friendly.

The RIFFLE clustering program looks for distinct sub-populations in the data. Generally, including more attributes in the data set gives RIFFLE more information to sift through, and results in a more complete clustering analysis. It is, consequently, preferable to group all of the attributes

describing a set of points together into a single file. This makes it possible to investigate the interaction and strength of all of the attributes together, with no loss in ability to investigate subsets of the attributes separately.

For example, two data sets may be recorded for each sample day in an experiment, one focusing on chemistry and the other on biological attributes. The two data sets describe the same points on the same day, but use disjoint sets of attributes for each description. While it may make sense to have two data sheets when collecting the data, for the clustering analysis it is better to join the data for each point making a single file that has all of the attributes.

Similarly, the RIFFLE program tends to work best when the groups are of the same size. A "control" group in the experiment should have the same number of points as the non-control groups. Likewise, all of the groups should have their attributes measured at identical intervals, and all of the groups should have a nearly identical composition (i.e. the same ratio of females to males). These considerations will improve the quality of the analysis by providing equal amounts of data to describe each group. It is a mistake to assume that fewer points are needed to describe the control group adequately, and such a strategy could undermine the experiment.

Guidelines

Data files will look something like the matrix in Figure 1 and follow the rules listed below.

```
1   5.1   3.5   1.4   0.2
1   4.9   3.0   1.7   0.2
1   4.7   3.2   1.3   0.4
⋮    ⋮     ⋮     ⋮     ⋮
2   4.5   3.8   1.0   0.3
⋮    ⋮     ⋮     ⋮     ⋮
3   5.2   3.0   1.8   0.5
⋮    ⋮     ⋮     ⋮     ⋮
```

Figure 1: Data file with three groups.

- The data file must be a rectangular matrix of numbers where the points (rows) are separated by a newline, and the attributes (columns) are separated by white space. For example, if there are 24 points and 37 attributes, then the file must have 24 lines, and each line must contain 37 numeric values. RIFFLE will accommodate missing values, but they must be coded as described below.
- All of the attributes describing the points should be joined into a single file. As mentioned above, a greater number of attributes gives more information for the analysis, while subsets of the attributes can easily be investigated.

- All values in the file must be numbers in either integer, real, or scientific notation format separated by white space.
- **All values in the data set must be greater than or equal to zero** except when describing a missing value. This may require that the numbers be translated so that the smallest value is zero or positive.
- Missing values must be be coded in the file as values $\leq$ -99. Typically we use -999 for missing values. This is done, in part, because this number is easy to spot in the data file.
- The data file cannot contain alphabetic characters (except when used in scientific notation) or special characters (except white space: space, tab, and newline).
- Attributes that are specifically identified as "discrete" for the RIFFLE program **must be coded as small consecutive integer values starting with 1**. Discrete in this context means that the set of values fall in distinct "bins" and the bins *do not* have an implicit ordering. For example, bins of "white", "brown", "black", and "spotted" may not have a relative ordering, whereas bins of "small", "medium", "large", and "extra-large" are implicitly ordered. This distinction may be important in some analyses. For instance, it may make sense to allow "small" and "medium" to cluster together, but not "small" and "extra-large". Whereas, it may be equally appropriate for "white" and "brown" to cluster together as it is for "white" and "spotted". In the latter case the attribute can be considered discrete.

  Discrete binary attributes, such as gender, **need not be coded as discrete**, instead, they can be coded as "continuous". Because RIFFLE is non-metric it makes no difference whether the continuous values are coded as 0 and 1, or 1 and 2, or 3 and 13, or 1.03 and 1.04. All of these binary codes will provide exactly two values that will differentiate the attribute's two possible states. Only if the attribute is identified to RIFFLE as "discrete" must the groups be 1 and 2.
- **Group identification values must be coded as consecutive integer values starting with 1. "Control" or non-dosed groups cannot be coded as 0.** For example dose groups which are originally coded by milligrams of toxicant would have to be translated to integer values starting at 1.
- A group identification attribute must be included for each point *unless* the groups are equal size and the points in each group are listed consecutively in the data file. In this case RIFFLE will assume (i.e. for n groups) that the first $\frac{1}{n}$ points are in one group, and the next $\frac{1}{n}$ points are in the next group, and so on.

Description File

In addition to the data file, RIFFLE will also accept a description file that provides labels for the attributes and identifies which attributes to include in the analysis. Figure 2 shows an example of such a description file.

- Each attribute name is a string without blanks (i.e. "Sepal_Length" with an underscore is O.K., as is "SepalLength", and "Sepal-Length", but "Sepal Length" with a space is not acceptable).

```
numberofgroups: 3
Group_Number    continuous   exclude   grouptag
Sepal_Length    continuous   include
Sepal_Width     continuous   include
Petal_Length    continuous   include
Petal_Width     discrete     include
```

Figure 2: Description file with five attributes, four included attributes, and three groups identified by the first attribute "`Group_Number`".

- Each name must appear on its own line.
- The order of the names must match the order of columns in the data file.
- The description file can optionally include the strings "include" and "exclude" following the attribute name on each line to tell RIFFLE which attributes to consider in the analysis. Data files often include some attributes that are not appropriate for statistical analysis, such as component attributes that are used to compute an aggregate attribute. Including the component attributes and the aggregate is redundant and inappropriate in the RIFFLE analysis. Either the components or the aggregate may be included, but not both.
- The description file can optionally include a string "grouptag" following one of the attribute names. The grouptag attribute identifies which group each point belongs to and is used in the association analysis to test for a statistically significant association between the known groups and the computed clusters. This attribute is automatically excluded from the clustering analysis, and is only used for the association statistic. However, it is good practice to explicitly exclude the grouptag attribute.
- The description file can optionally include a string "numberofgroups: n" where n is the number of groups in the file. This feature can be used to split the points into equal sized groups when there is no "grouptag" attribute. This can be useful when there is no grouptag attribute, when it is desirable to divide the points into n groups for one analysis and then into m groups for another analysis, or when the number of groups does not equal the number of clusters sought.

Data files should be accompanied either by description files that have the above information or a written description that gives instructions on these issues.

Appendix 4

Selected Reprints Using Nonmetric Clustering

Ecological Modelling, 53 (1991) 167–187
Elsevier Science Publishers B.V., Amsterdam

Classification and ordination of limnological data: a comparison of analytical tools

R.A. Matthews [a], G.B. Matthews [b] and W.J. Ehinger [c]

[a] *Huxley College of Environmental Studies,*
[b] *Computer Science Department,*
Western Washington University, Bellingham, WA 98225, U.S.A.
[c] *Department of Biology, University of North Carolina, Chapel Hill, NC 27514, U.S.A.*

(Accepted 7 May 1990)

ABSTRACT

Matthews, R.A., Matthews, G.B. and Ehinger, W.J., 1991. Classification and ordination of limnological data: a comparison of analytical tools. *Ecol. Modelling*, 53: 167–187.

In this paper we compare the differences between principal components analysis, hierarchical clustering, correspondence analysis and conceptual clustering to show their effectiveness for identifying patterns in a large limnological data set. The data for this comparison come from a multi-year study of Lake Whatcom, a large lake located in the Puget Sound lowlands of the state of Washington. The data include both physical and chemical parameters (temperature, dissolved oxygen, pH, alkalinity, turbidity, conductivity, and nutrients) as well as biological parameters (Secchi depth, chlorophyll *a*, and phytoplankton species and total counts). The patterns we expected to find include (a) temperature and dissolved oxygen interactions, (b) ordination by algal bloom sequences, and (c) clustering due to the effects of stratification.

Principal components analysis was somewhat useful for confirming known water quality trends, but did not successfully identify large-scale patterns such as stratification and seasonal plankton changes. Correspondence analysis proved to be superior to principal components analysis for detecting phytoplankton trends, but was not as good for interpreting water quality changes. Hierarchical clustering produced highly unbalanced trees for both the water quality and phytoplankton data, and was useless as an exploratory tool. A new approach to clustering, implemented in the computer program RIFFLE, is introduced here. This clustering algorithm outperformed the other exploratory tools in clustering and parameter ordination, and successfully identified a number of expected and unexpected patterns in the limnological data.

INTRODUCTION

One of the most difficult problems in aquatic ecology is the interpretation and modelling of the complex data sets that are generated from limnological

0304-3800/91/$03.50

research. The data generally are not linear, rarely conform to parametric assumptions, and are often measured using incommensurable units such as length, concentration, and frequency. In addition, most limnological research generates incomplete data sets, not only because of sample loss, but also due to sampling design. For example, lake depth, temperature, and dissolved oxygen may be measured every few meters from the surface to the bottom, while plankton populations are usually sampled only in the photic zone. As a result, we may have to rely on the robustness of a statistical test to identify significant trends despite violation of the test's fundamental assumptions. Further, true gradients, as understood in terrestrial ecology, are rarely present. Nevertheless, patterns of algal blooms and successions are present, and their recognition poses an important problem for data analysis and modelling.

In this paper we compare several types of analytical procedures, including graphical analysis, hierarchical clustering, and ordination (principal components analysis and correspondence analysis), to see how well they identify patterns in a large limnological data set. While all of these methods are in common use, they are not all equally useful for identifying patterns in ecological data sets (Pielou, 1984; Ludwig and Reynolds, 1988). In addition, we used a new version of conceptual clustering (Fisher and Langley, 1986), which turned out to be markedly superior to correspondence analysis in parameter ordination, and superior to hierarchical techniques in clustering.

Our data come from Lake Whatcom, a large monomictic lake in Washington. Water quality data have been collected from Lake Whatcom since the early 1960's, with intensive sampling since 1982. The data for this paper are from spring 1987 through winter 1988 because this period included intensive plankton sampling as well as water quality monitoring. The patterns we expected to find in the lake included: (a) temperature and dissolved oxygen interactions, (b) algal bloom sequences, and (c) indicators and effects of stratification. Evidence for all of these was discovered in the data set. However, some of the analytical techniques were less useful than others for identifying the limnological trends. We have included a general discussion of the fundamental differences between each analytical technique as well as a summary of the strengths and weaknesses of each technique for identifying patterns in limnological data.

METHODS

Study site

Lake Whatcom is a 2000 ha chain lake located in the Puget Sound lowlands of northwestern Washington (Fig. 1). The lake is divided into three

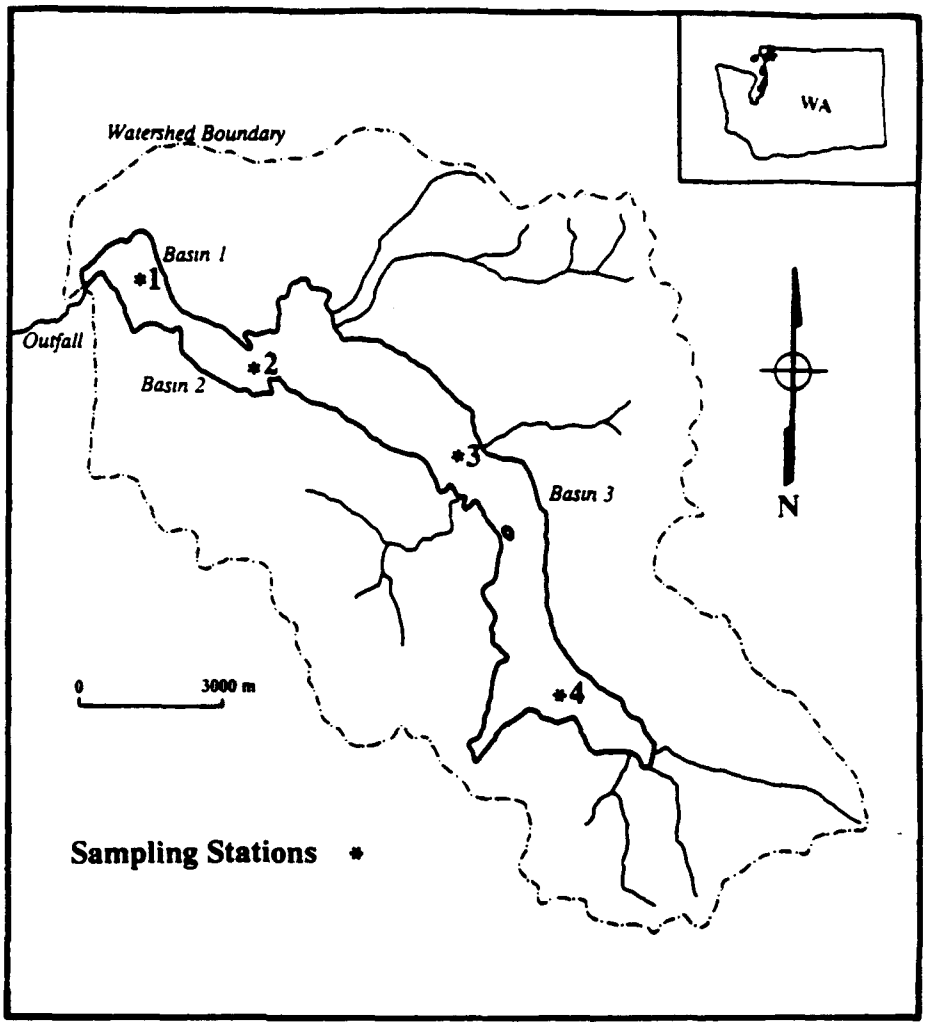


Fig. 1. Lake Whatcom sampling sites, Whatcom County, Washington.

distinct basins by subsurface sills; the largest basin, Basin 3, contains 96% of the lake volume, while Basins 1 and 2 each contain about 2% of the total lake volume (Lighthart et al., 1972). Lake Whatcom is a warm, monomictic lake; the direction of flow is from Basin 3 → Basin 2 → Basin 1. All of the perennial streams in the Lake Whatcom watershed drain into Basin 3. The only natural outflow from the lake is Whatcom Creek in Basin 1. However, the city of Bellingham withdraws water from Basin 2 for municipal drinking water and industrial uses. In the summertime the municipal withdrawal is often the only significant outflow from the lake.

Water quality and phytoplankton sampling

Water samples were collected at four sites in Lake Whatcom (Fig. 1) from March 1987 to October 1988. Temperature, pH, conductivity, and dissolved oxygen were measured in the field using a Hydrolab Surveyor II. In Basins 1 and 2, where the maximum depths are 20 and 22 m, respectively, these measurements were taken at 2-m intervals from the surface to the bottom of the water column. In Basin 3 (maximum depth > 90 m), the measurements were taken at 2-m intervals to the depth of 20 m, and at 5-m intervals from 20 m to the bottom. Secchi depth was also measured in the field at each site.

The water samples for nutrients analyses (ammonia, nitrate/nitrite, total nitrogen, soluble reactive phosphate, and total phosphorus), total organic carbon, and dissolved inorganic carbon analyses were collected at 5-m intervals in Basins 1 and 2, and 10-m intervals in Basin 3. The nutrient analyses were done using a Technicon Autoanalyzer, following EPA (1983) guidelines for sampling handling and analysis. The total organic carbon and dissolved inorganic carbon analyses were done using an OIC Model 0524B Infrared Carbon Analyzer (APHA, 1985).

All chlorophyll and phytoplankton samples were collected at 5-m intervals from the surface to 15 m (phytoplankton) or 20 m (chlorophyll). Chlorophyll *a* extractions were done by filtering 250–500 mL of sample through a glass fiber filter, which was ground in a tissue grinder and extracted with 90% spectrophotometric grade acetone. The chlorophyll *a* concentrations, corrected for phaeophytin *a*, were measured using a calibrated Turner Designs fluorometer (APHA, 1985). Phytoplankton samples were preserved with Lugol's solution, and were identified and counted using a Sedgewick-Rafter counting chamber on an Olympus Inverted Microscope (APHA, 1985; Lind, 1985). Representative phytoplankton samples were sent to the Academy of Natural Sciences of Philadelphia for taxonomic verification.

Data analysis methods

The data were analyzed using either ordination, clustering, or both. Ordination of 'points' (all measurements collected at a particular date, site, and depth, sometimes called 'samples' or 'sampling units') was done by principal components and correspondence analysis (reciprocal averaging). Ordination of 'parameters' (e.g., pH, temperature, etc., sometimes called 'attributes", 'dimensions', or 'variables') was done by correspondence analysis and conceptual clustering. Clustering was done with an agglomerative, hierarchical algorithm, as well as with an optimizing, conceptual clustering

algorithm. Visual confirmation of patterns in the data was made using two- and three-dimensional graphical displays of the data.

Point ordination

Principal components analysis was done using data normalized by mean and standard deviation (z-scores), using the FACTOR procedure provided in the SPSS-X statistical package. This resulted in several ordinations of the points, one for each principal component. Generally, the first three or four principal components were inspected graphically.

Correspondence analysis (reciprocal averaging), which simultaneously ordinates both the parameters and the data points, has proven better than principal components analysis in the analysis of many kinds of ecological data. In data sets involving large-scale gradients in the environment, for example, with high beta diversity along the gradients, correspondence analysis outperforms principal components analysis (Kenkel and Orloci, 1986). It can be used for detecting unknown gradients or confirming the existence of expected ones. Correspondence analysis scores were computed directly using the iterative technique (Pielou, 1984, pp. 184–188).

Hierarchical clustering

Hierarchical clustering uses a measure of similarity or distance between points, and derived measures of inter-cluster and intra-cluster distance. It is hierarchical in that each cluster is a subcluster of a larger cluster; the total clustering forms a tree, or dendrogram. Balanced dendrograms indicate a good clustering into roughly equal-sized clusters, while unbalanced dendrograms indicate little real clustering, but instead a gradual agglomeration of sample points into a single group.

The choice of a distance measure is often critical to hierarchical clustering (Ludwig and Reynolds, 1988). We employed two distance measures for hierarchical clustering: squared Euclidean distance, defined as $\Sigma_i(x_i - y_i)^2$ and cosine of vectors distance, defined as $\Sigma_i(x_i y_i)/\sqrt{(\Sigma_i x_i^2)(\Sigma_i y_i^2)}$, where x_i and y_i are the parameter values for two points. Cosine distance is similar to chord distance (Ludwig and Reynolds, 1988), and considers only the relative proportions of the various parameters that make up a sample point. Squared Euclidean distance also takes into account the absolute size of parameter values.

The algorithm we used for forming the hierarchy of clusters was average linkage between clusters. This method gives good results on synthetic, Gaussian data known to have well-defined clusters (Bayne et al., 1980).

Conceptual clustering

The philosophical difficulty with hierarchical clustering is that it assumes the meaningfulness of combinations of parameters, such as the Euclidean and cosine distances, above. In ecological data sets, such compositions as these two are often not meaningful, due to incommensurability. For example, an uncommon organism with a large individual biovolume may have the same total biomass as a common organism with a smaller individual biovolume, but since both species are measured in organisms per L, the common organism will dominate in terms of absolute number and proportion. Predators, for example, often fall into this category, being generally large in size but small in number. However, their functional importance would be overlooked by this analytical technique which would simply add or multiply the two numbers. The problem lies not in the manner of counting organisms, but in the necessity to *combine* counts of dissimilar species. The problem is even worse for water quality data, where different parameters are measured in degrees, pH units, concentrations, and so on.

Conceptual clustering can be used as an alternative to hierarchical clustering [see Fisher and Langley (1986) for a survey]. A clustering technique is called 'conceptual' if it yields descriptions of the clusters in terms of *concepts*, i.e., in terms of only conceptually important parameters. What is 'conceptually important' depends on context, but in scientific data analysis we take the following as an acting principle: Clusters are conceptually important if knowledge of such clusters increases the reliability of predictions about parameter values. In other words, we seek clusters such that most (if not all) of the actual observed data values for a sample can be predicted more accurately after its cluster has been identified than before such identification. Thus, 'conceptually important' clusters, in our methodology, are those that warrant accurate predictions of parameter values.

We developed a clustering tool, called RIFFLE, in line with these principles, which is superior to traditional clustering methods for a wide range of ecological data sets (Matthews and Hearne, 1991). A brief description of the algorithm is given in Appendix A. RIFFLE has the following advantages over traditional clustering methods: (1) Measures based on combinations of incommensurable parameters, such as Euclidean distance in parameter space are not used, (2) transformations of scale do not affect the outcome, (3) parameters can be nominal, ordered, numeric, or mixed, (4) 'noisy' parameters, i.e., those with large variance but little association with any other parameters, are automatically filtered out and have little effect on the resulting clustering, (5) 'rare' parameters, i.e., those with small variance but with a significant correlation to the dominant patterns of the data set, are automatically given weight in accord with that correlation, and (6) no

assumptions about points with missing values, such as replacement with zeroes or with the mean, need to be made. RIFFLE simultaneously clusters the data and ordinates the parameters in terms of their conceptual significance to the clusters. It is thus, in a sense, similar to correspondence analysis in that simultaneous analysis of points and parameters is done, except that a *non-linear* patterning the points (a clustering) is sought together with a *linear* ordination of the parameters. Correspondence analysis attempts to provide a linear ordination of both.

RESULTS AND DISCUSSION

Physical-chemical data

The physical-chemical data from Lake Whatcom indicate that the three basins are dissimilar, which is best illustrated by comparing graphs of the temperature and dissolved oxygen data for the four sites (Figs. 2–5). The two shallow basins (Basin 1, Site 1 and Basin 2, Site 2) both had significant oxygen deficits, and both developed anoxic hypolimnia during the summer. Basin 3, Site 3, experienced some oxygen depletion during the summer; however, the oxygen concentrations usually did not fall below 2 mg/L. Basin 3, Site 4 maintained consistently high dissolved oxygen levels throughout summer stratification, even at the bottom of the water column.

The oxygen deficit in Basin 1 was more pronounced than in Basin 2. This observation was discussed by Ehinger (1988) and is thought to be due, at least in part, to isolation of Basin 1 during the summer when the outflow from the lake into Whatcom Creek was reduced to near zero. The City of Bellingham continued to withdraw water from Basin 2 throughout the summer, which flushed Basin 2 with high quality water from Basin 3.

The remaining water quality parameters were strongly influenced by the temperature and dissolved oxygen conditions in the lake. Basins 1 and 2 experienced epilimnetic nitrate depletion during summer algal blooms. Concurrently, ammonia and phosphate were released from the sediments and accumulated in the hypolimnia of both basins. In Basin 3, similar conditions developed, but to a much lesser extent. Alkalinity and pH values showed little variation except during stratification. During this time, the pH values were slightly higher in the epilimnia of Basins 1 and 2 due to photosynthetic activity, while the pH values in the hypolimnia were lower due to the release of reduced compounds from the sediments. Similarly, the alkalinity values increased slightly near the sediments during stratification. Conductivity, turbidity, dissolved inorganic carbon, and total organic carbon values were fairly uniform throughout the sampling period. A complete listing of the

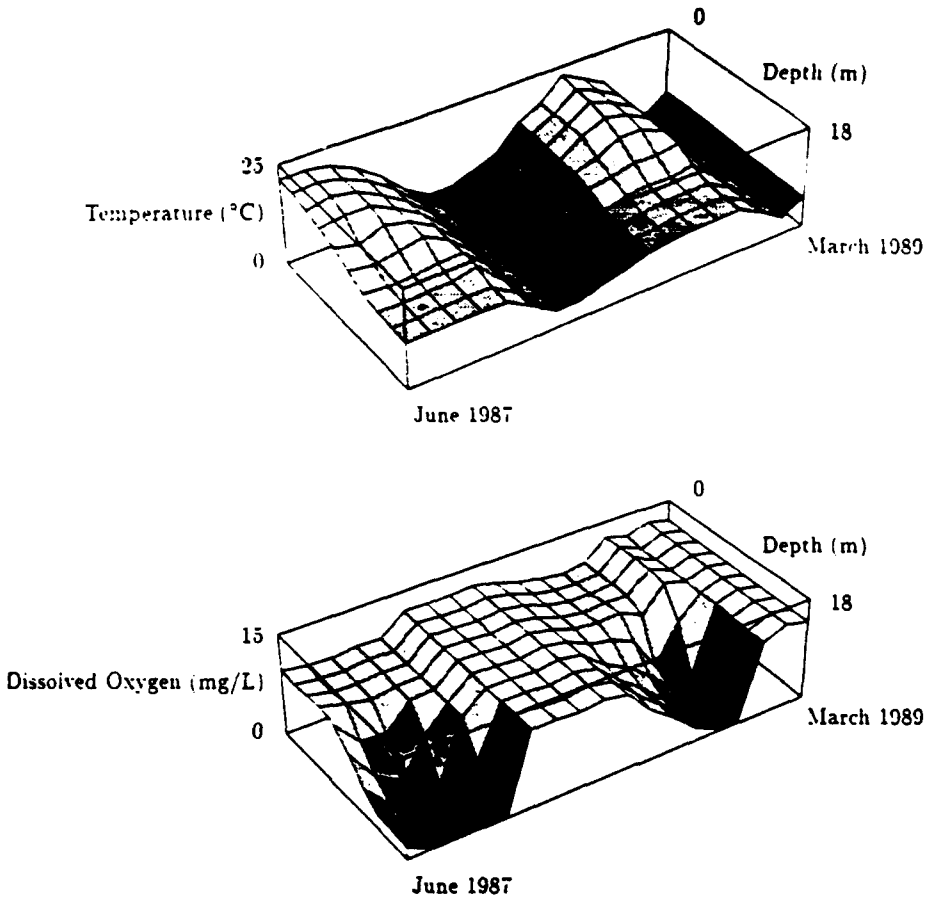


Fig. 2. Temperature and dissolved oxygen profiles for Basin 1, Site 1.

water quality data is available from the authors, and a list of parameters sampled is in Appendix B.

Conceptual clustering of the physical-chemical data proved to be best at confirming the expected trends. Figure 6 shows how RIFFLE clustered the physical and chemical data for each discrete sample set (matched by date, site, and depth class). The RIFFLE clusters were plotted by the date and temperature value for the data set so that the influences of thermal stratification can be observed. Sample points were grouped into classes based on approximate ($\approx$ 5 meter) depth, and data values were taken as averages of the values in a single depth class. Depth classes were used because of the large number of points in the Hydrolab data sets ($>$ 1600 for each parameter) and because there was some variation in the depth of some samples. For example, the 'bottom' measurements varied by several meters, depending on

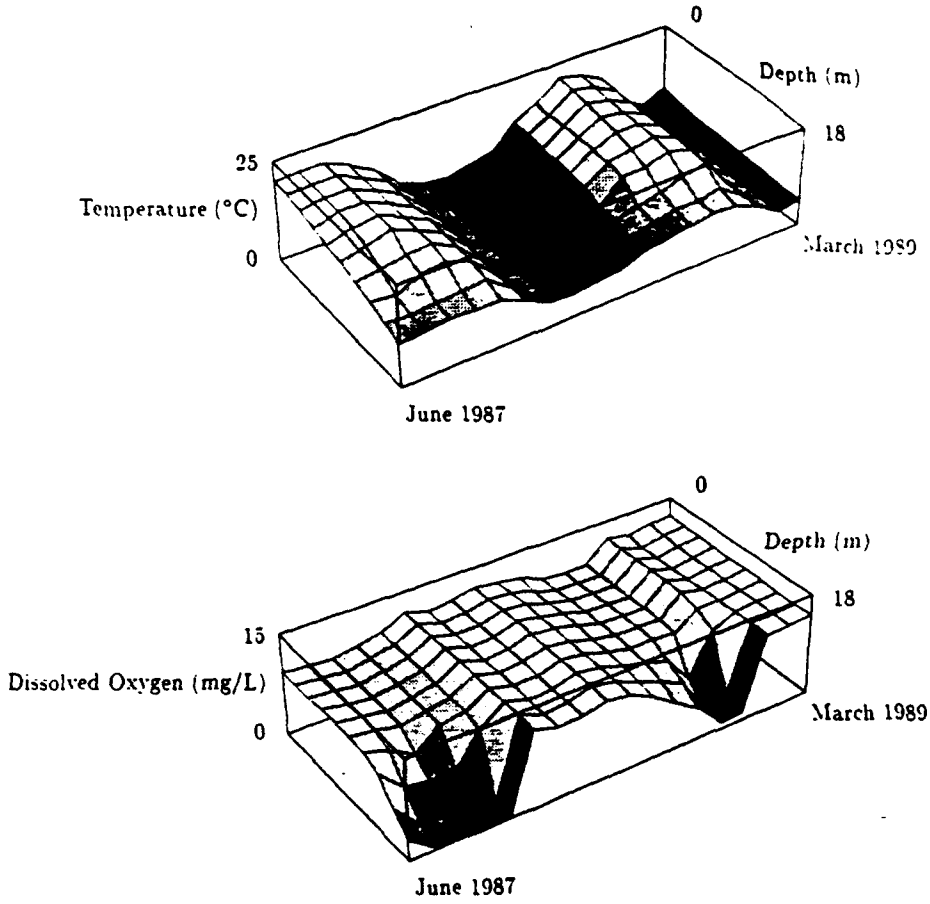


Fig. 3. Temperature and dissolved oxygen profiles for basin 2, Site 2.

where the boat was located. A smaller total number of points also helped in the graphical presentation of the data.

In Basin 1, three clusters were selected as best describing the data. Two of the clusters (○ and ▷) separate the epilimnion and hypolimnion samples during stratification, while the third cluster (★) identifies the well-mixed samples of the unstratified period. The vertical lines marking stratification and turnover were estimated from the temperature data for each basin; however, the exact timing of these events was not determined. This is important because most of the misclassifications in the RIFFLE clusters occurred within one sampling date of our estimated dates for stratification or turnover.

Basin 3 clustered into only two groups: stratified epilimnial samples (▫) and a second group consisting of both hypolimnial samples and mixed lake

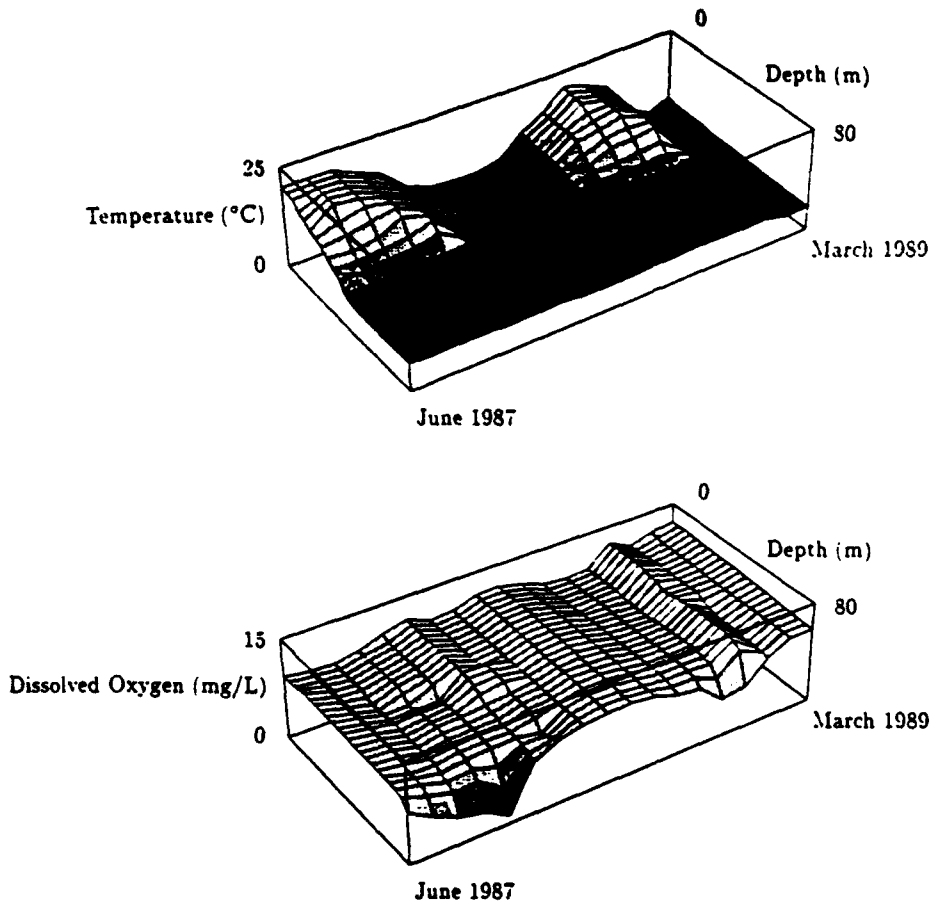


Fig. 4. Temperature and dissolved oxygen profiles for Basin 3, Site 3.

samples (★). This supports our temperature and dissolved oxygen data that show Basin 3 to be oligotrophic, with little change in the hypolimnetic water quality occurring during summer stratification.

In Basin 2, a unexpected pattern emerged. During stratified periods, three clusters were identified. Upon closer inspection of the temperature and dissolved oxygen data, we found that the depth of the thermocline was deeper in Basin 2 than in Basin 1, and the height of the anoxic portion of the hypolimnion (0–2 mg/L) was much higher in Basin 1 than in Basin 2. In Basin 1, both the surface and the 10-m depth classes would lie primarily in the epilimnion, while the remaining measurements (20 m and bottom) would be in the hypolimnion, and strongly influenced by anoxic conditions. However, in Basin 2, the 10-m depth class would be at the thermocline and slightly above the anoxic portion of the hypolimnion. The remaining sam-

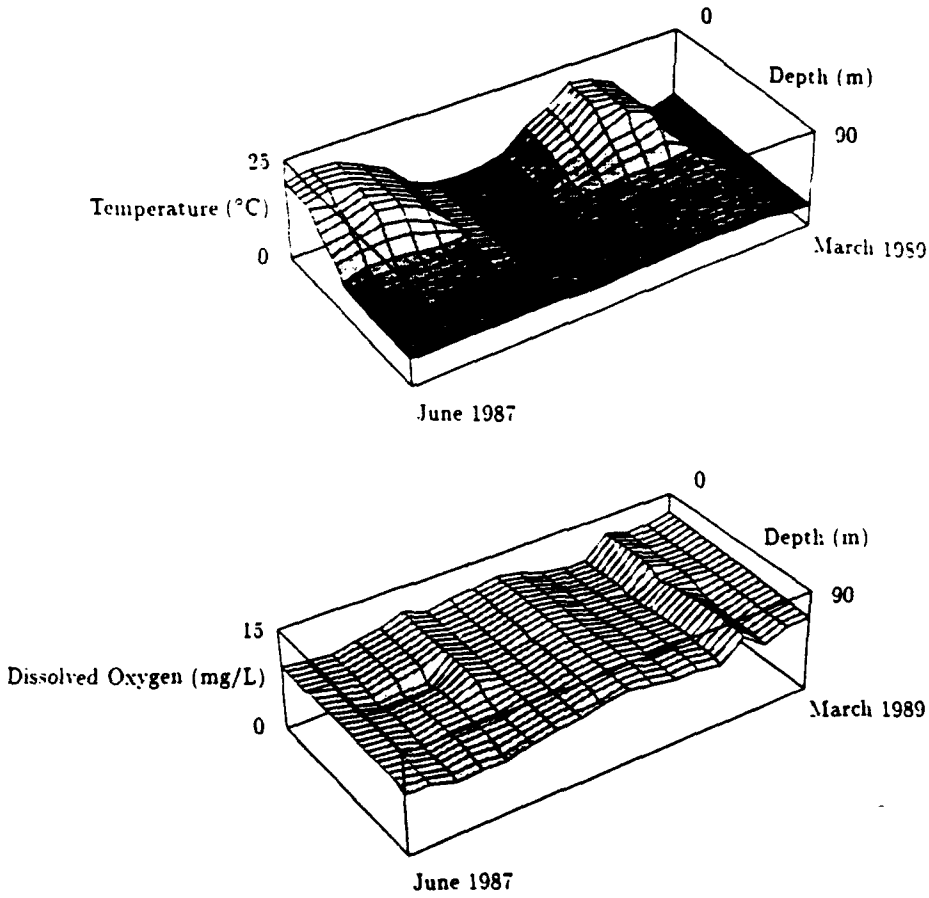


Fig. 5. Temperature and dissolved oxygen profiles for Basin 3, Site 4.

ples (at 20 m and below) would reflect hypolimnetic influences. The three clusters in Basin 2, therefore, identify the epilimnion, metalimnion, and hypolimnion.

Principal components analysis did not work well when plotted by individual basins, but did identify the major trends for the entire data set; The first principal component accounted for 24% of the total variance; its dominant terms (with a factor greater than 0.5) were:

$$0.872\ \text{Temperature} - 0.842\ \text{Depth} + 0.735\ \text{pH} - 0.623\ \text{Nitrate/Nitrite}$$

The second principal component accounted for another 19% of the total variance and its dominant terms were:

$$-0.779\ \text{Dissolved Oxygen} + 0.695\ \text{Turbidity} + 0.663\ \text{Alkalinity}$$

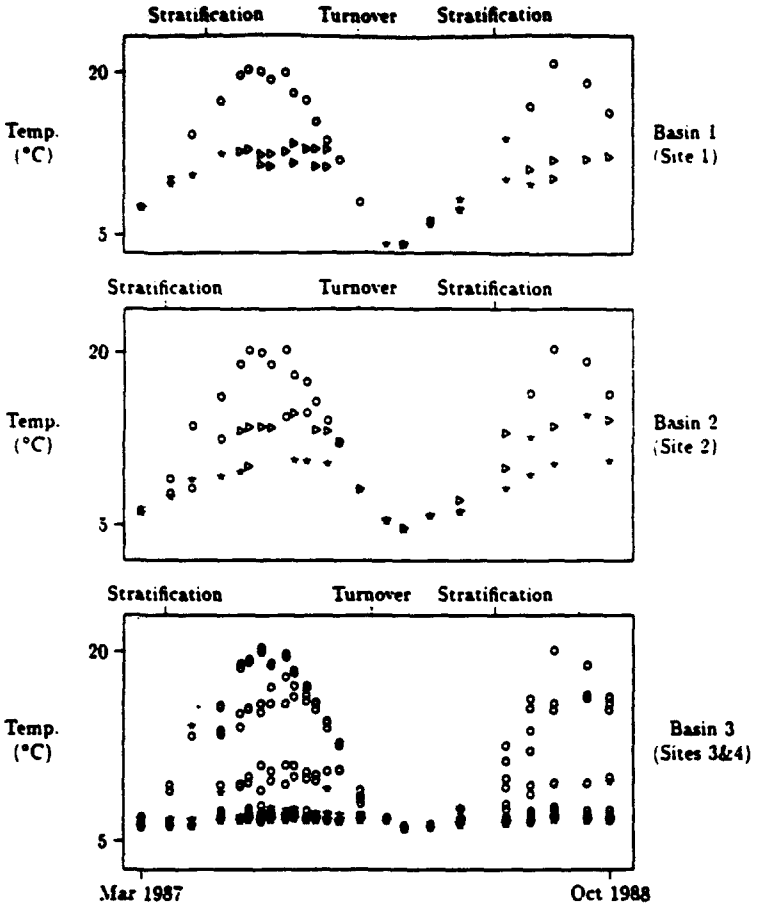


Fig. 6. RIFFLE clustering of chemical data. Conceptual clusters (○, ▷, and ★) plotted by temperature and date.

The first principal component identified the inverse relationship between temperature and depth during summer stratification as well as the changes in pH and nitrate values that were discussed earlier for Basins 1 and 2. The second component picked up on the hypolimnetic oxygen depletion that was observed, to a greater or lesser extent, in all three basins following stratification. The positive turbidity factor was probably an artifice that resulted from sampling too near the sediments, while the alkalinity factor again reflects the effects of biological activity during stratification.

Hierarchical clustering and correspondence analysis did not identify any meaningful trends in this data set. Correspondence analysis found nearly all points to have the same scores, and thus any parameter ordination was of doubtful validity. Hierarchical clustering resulted in unbalanced dendro-

grams, and had the added disadvantage that, since points with missing data could not be included, the data had to be severely subsetted. Several parameters (Secchi depth, dissolved inorganic carbon, and total organic carbon) had to be excluded because they were measured less frequently than other parameters.

Phytoplankton data set

Since it is only useful to collect phytoplankton data at or near the surface, this data set is considerably smaller, in terms of number of points, than the physical-chemical data set. A complete listing of taxa found is provided in Appendix C.

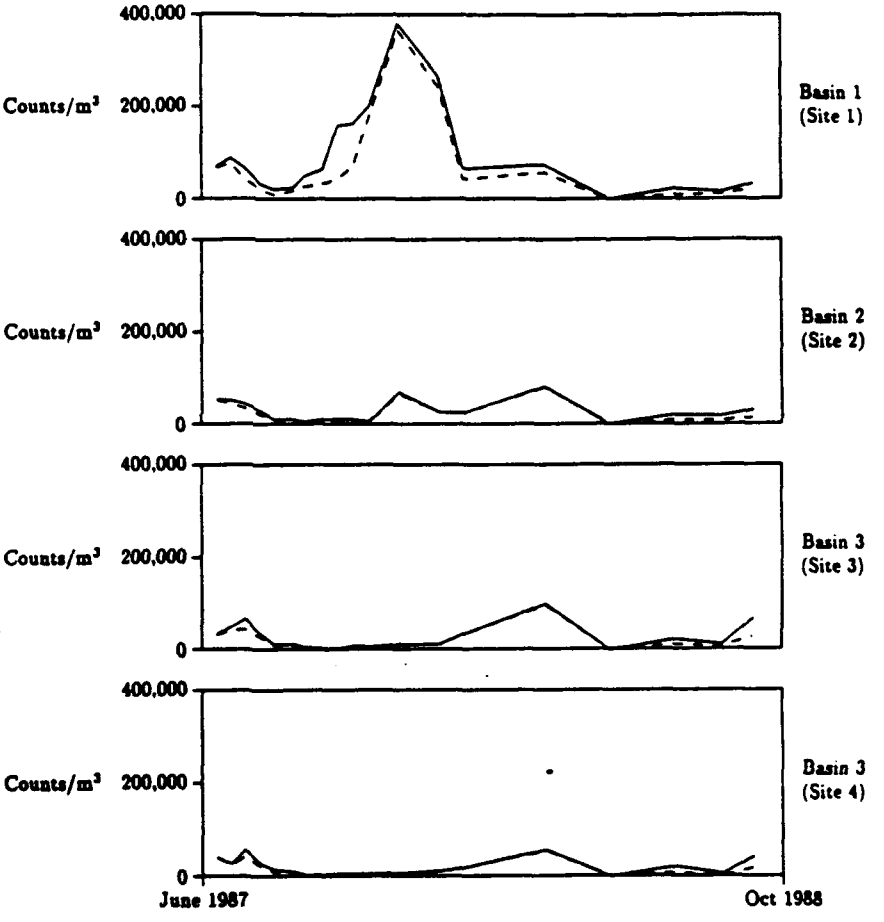


Fig. 7. Total phytoplankton (solid) and diatoms (dashed) in Lake Whatcom.

Figure 7 shows a summary of the phytoplankton data for Lake Whatcom. Diatoms (predominantly *Melosira ambigua* (Grun.) O. Mull, *Melosira distans* (Ehr.) Bethge, and *Fragilaria crotonensis* (Kitt.) dominated the phytoplankton populations most of the year, with peaks occurring during the winter and spring.

During the late summer (during periods of nutrient depletion in the epilimnion), blooms of mostly green and bluegreen algae developed, especially in Basin 1. The densities of green and bluegreen algae never reached the peak densities that were measured for the winter/spring diatom blooms. This is partly due to our system of counting, whereby *Coelosphaerium naegelianum* Unger, a common late summer bluegreen alga, was counted by colonies rather than individual cells. If *Coelosphaerium* had been counted by individual cells (not an easy task) or if each plankton count was weighted to account for biovolume, [as in Ehinger (1988)], the *Coelosphaerium* total

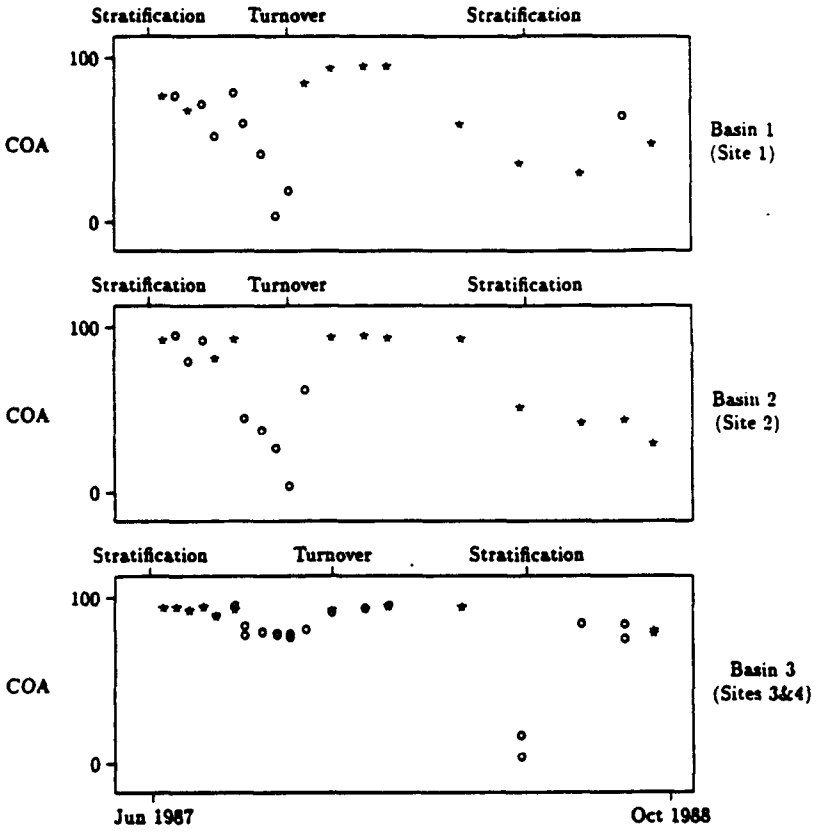


Fig. 8. RIFFLE clustering of phytoplankton data. Conceptual Clusters (○ and ★) plotted by correspondence analysis score and date.

'count' would increase. This problem of counting individuals, colonies, or biovolumes is frequently encountered in limnological data sets, and is part of the reason why the statistical tool needs to be insensitive to scale.

Conceptual clustering of the phytoplankton data again proved valuable for identifying the major trends in the lake. Figure 8 shows the clusters generated by RIFFLE plotted by correspondence analysis score (COA) vs. time. In all three basins, samples collected before and after turnover tended to be in different clusters; similar rapid changes in the phytoplankton populations did not occur following stratification. Turnover is a dramatic event in lakes, often occurring within a few days, that causes rapid changes in the water quality of the lake. Stratification, however, causes a gradual divergence of the water quality in the epilimnion and hypolimnion. In Fig. 8, late summer phytoplankton (○) were clearly distinguished from post-turnover phytoplankton (★). However, the late summer phytoplankton populations were not reestablished until several months after the onset of stratification.

In creating the clusters shown in Fig. 8, RIFFLE clustered temporally adjacent points. This is in line with the proposed existence of temporal 'plateaus' in phytoplankton succession, mentioned in (Legendre et al., 1985). RIFFLE, however, clustered them successfully without the ad hoc imposition of an explicit chronological constrait or the elimination of singleton clusters.

The taxa that contributed most heavily to the RIFFLE clusters included many common species (e.g *Fragilaria* and *Coelosphaerium*), but also included several 'rare' species that were highly correlated with turnover. One example is *Ceratium hirudinella* (O.F. Muell.), a large dinoflagellate, that never occurred in large numbers, but was only collected during late summer just prior to turnover. *Ceratium* is able to compete well during late summer because it can swim to positions of optimum light and nutrient concentrations. Because of its low density in Lake Whatcom, none of the other statistical tools used *Ceratium* to identify late summer phytoplankton blooms. RIFFLE's ability to use both common and rare taxa is particularly useful for finding potential indicator species.

Principal components analysis was able to identify the major phytoplankton blooms; however, the results could easily be misinterpreted if importance was assigned to the individual species comprising each principal component rather than the trend that those species represent. For example, the winter diatom bloom was represented by *Melosira*, *Fragilaria* and *Tabellaria flocculosa* (Roth) Kutz. in the combined data set, but only by *Fragilaria* and *Melosira* in Basin 1 (see Table 1). This does not mean that *Tabellaria* was absent or rare in Basins 2 and 3; only that it accounted for less variation in the data sets for those basins. The interpretation of the summer phytoplankton blooms is even more difficult: the representative species are split into two groups in Basin 1, but only one group in the

TABLE 1

Principal components for Lake Whatcom phytoplankton, Basin 1 and all basins combined

Basin 1	Total	Species	Loading
PC-1	23%	*Dictyosphaerium* sp.	0.94
		Staurastrum sp.	0.93
		Aphanocapsa sp.	0.93
PC-2	16%	*Rhabdoderma* sp.	0.89
		Chroococcus sp.	0.87
		Oscillatoria sp.	0.85
PC-3	11%	*Fragilaria crotonensis*	0.93
		Melosira sp.	0.93
All Basins	Total	Species	Loading
PC-1	15%	*Dinobryon* sp.	0.790
		Coelsphaerium naegelianum	0.769
		Eudorina elegans Ehrenberg	0.774
		Unknown Greens	0.667
		Aphanocapsa sp.	0.542
PC-2	10%	*Melosira* sp.	0.905
		Fragilaria crotonensis	0.854
		Tabellaria flocculosa	0.847

combined data, and there is little overlap between the species in the different groups. While in some cases these results might lead to the discovery of an unknown pattern in the data, close inspection of the Lake Whatcom data does not support any such conclusion.

Correspondence analysis was more revealing. As can be seen from Fig. 8, there is a tendency for the COA score gradually to lessen during stratification, and swing rapidly back to its highest values immediately following turnover. This indicates that the large-scale gradient from a mixed to a stratified lake can be detected by correspondence analysis, and that the Lake Whatcom sample points successfully ordinated according to this trend. Basin 3, however, reveals that the presence of outliers can have a disastrous effect on this ordination technique. Gauch et al. (1977) make the same observation.

Hierarchical clustering proved ineffective in handling the Lake Whatcom phytoplankton data, typically resulting in highly unbalanced trees, whether squared Euclidean distance or cosine distance was used. The tree development was disastrously affected by outliers. Modification can be made to hierarchical clustering that improve its use for chronological samples. These modifications iclude: (a) transformations of the data matrix (normalization etc.), (b) the explicit removal of outliers from the data set during clustering, and (c) the imposition of a constraint to force temporally adjacent sample

points into the same clusters [see Allen et al. (1977); Legendre et al. (1985)]. However, these constraints seem excessively severe to us, and conceptual clustering provides an excellent alternative.

CONCLUSIONS

We conclude that limnological data sets are amenable to clustering and gradient analysis, with the proviso that care must be taken in the tools used. Principal components analysis was of some use in confirming water quality trends, in that it achieved a reduction in the redundancy of the data set by combining correlated parameters (such as temperature and pH) into a single component. However, principal components did not aid in the identification of large-scale patterns in the data, such as stratification. Further, used on data sets with many parameters (such as species lists) principal components provided only a marginal reduction in the complexity of the raw data. We found correspondence analysis to be superior to principal components for detecting large-scale gradients in the phytoplankton data from Lake Whatcom. This is consistent with the findings from theoretical studies of ordination (Kenkel and Orloci, 1986).

We believe that the results of this study, in conjunction with similar studies at other sites, will lead to an improvement in conventional biogeochemical modelling of limnological systems. Typically these models are lumped-parameter conceptual models, involving two crucial tasks. First, the model must be built on a small number of significant components, e.g. phosphorous, chlorophyll, phytoplankton or zooplankton, and, second, the gross, qualitative behavior of the lake must be understood in terms of changes in the states of these components (Scavia and Robertson, 1979, pp. 1–83). Conceptual clustering by RIFFLE helps by providing objective leads in both of these tasks: It provides an estimate, for each parameter, of how strongly the entire system is associated with that parameter; these estimates can guide the selection of components. It also provides a clustering of the samples of the lake system into states that may be significant parts of the evolution of the model.

Conceptual clustering was found to be consistently superior to hierarchical clustering. In clustering the physical chemical data, the presence of epilimnion and hypolimnion was clearly confirmed by our conceptual clustering algorithm. Hierarchical clustering did not isolate these clusters. In the phytoplankton set, a division into mixed and stratified communities was accomplished only by the conceptual clustering algorithm. This, together with the facts that (a) conceptual clustering makes fewer assumptions about the data than hierarchical clustering, and (b) it can handle incomplete and

mixed data sets without further assumptions or data subsetting, makes it a consistently superior tool for clustering.

ACKNOWLEDGEMENTS

We would like to thank the Institute for Watershed Studies, Western Washington University, and the City of Bellingham for their assistance with this project.

APPENDIX A

RIFFLE clustering

Clustering by the RIFFLE program (Matthews and Hearne, 1991) is a technique especially adapted to clustering ecological data. It is a partitional clustering algorithm: the data points are partitioned into clusters in a variety of ways, and the best such partition is selected as an appropriate clustering for the data. The 'best' clustering is one which maximizes the value of a fitness-measure (which evaluates the 'fitness' of the clusters to the data). The fitness-function used in RIFFLE estimates the accuracy of predictions in an imagined experiment, an experiment that uses the proposed cluster-membership of a sample to 'predict' whether that sample will have large or small values on its measured parameters. If a large number of these 'predictions' agree with the actual sample values, then the clustering fits. We use a nonparametric measure of fitness in the sense that predictions of numeric parameters are limited to the coarseness of the clustering. In a clustering into two groups, for example, only two values are predicted: 'high' values and 'low' values.

The quantitative measure of prediction accuracy used in RIFFLE is the proportional reduction in error, or *Guttman's* λ (Goodman and Kruskal, 1954). Suppose we wish to measure the fitness of a clustering into two groups, and we want to measure the accuracy of prediction for, say, a taxon t. Let a data point be represented by the vector x, with the point's value on parameter t be x_t. Let the two clusters be denoted by k_1 and k_2, and, for taxon t, let t_1 denote a 'high' value, and t_2 denote a 'low' value. (The best split value between 'high' and 'low' is also determined by the RIFFLE algorithm, but for concreteness we can assume the median is used.) A two-dimensional cross-tabulated frequency table, F, of the joint probabilities, is then built, where

$$F_{i,j} = \left| \{ x: x \in k_i \quad \text{and} \quad x_t = t_j \} \right|$$

i.e., F_{ij} is the number of times a sample is found which is in the ith cluster and has the jth value (high or low) of the taxa.

Under the usual statistical assumption that the distribution of sample points in F is representative of the distribution in the population, we can use F, and a knowledge of a sample's cluster, to predict the taxa count for that sample. If our sample is in cluster k_2, for example, our guess will be 'high' or 'low' depending on whether F_{21} or F_{22} has the larger value, and similarly if our sample is in cluster k_1.

If we do this for many samples, our total fraction of correct guesses C can be estimated to be:

$$C = \frac{\sum_i \text{Max}_j F_{ij}}{N}$$

where N is the total number of samples. The fraction on which we will be in error, then, will be $1 - C$. On the other hand, without a knowledge of a sample's cluster (and without using F), we can do no better in predicting 'high' or 'low' than 50% correct, on average (assuming a median split value). Our proportional reduction in error, therefore, using this clustering and its cross-classification table F, will be estimated to be:

$$\frac{(\text{Random Error}) - (\text{Clustered Error})}{\text{Random Error}} = \frac{1/2 - (1 - C)}{1/2} = 2C - 1$$

The RIFFLE program searches over a large number of partitions of the data in order to maximize this proportional reduction in error for a large number of measured parameters. In other words, it searches for the one clustering (out of many) which is most closely associated with the measured parameters.

This algorithm has been implemented in Pascal and has been tested on a wide variety of computers and data sets (Matthews and Hearne, 1991).

APPENDIX B

Lake Whatcom water chemistry parameters sampled

Temperature	pH
Conductivity	Dissolved oxygen
Turbidity	Alkalinity
Secchi disk	Ammonia
Nitrate/Nitrite	Total nitrogen
Soluble reactive phosphate	Total phosphorus
Total organic carbon	Dissolved inorganic carbon
Chlorophyll *a*	

APPENDIX C

Lake Whatcom phytoplankton taxa list

Phylum: Chrysophyta	
Anomoeoneis serians (Breb. ex Kutz)	*Asterionella formosa* Hass.
Cyclotella compta (Ehr.) Kutz.	*Dinobryon* sp.
Fragilaria crotonensis Kitt.	*Melosira ambigua* (Grun.) o. Mull.
Melosira distans (Ehr.) Bethge.	*Navicula* sp.
Stephanodiscus sp.	*Synedra chaseana* (Thomas) Boyer
Synura sp.	*Tabellaria flocculosa* (Roth) Kutz
Phylum: Cyanophyta	
Anabaena sp.	*Anacystis* sp.
Aphanocapsa sp.	*Chroococcus* sp.
Coelosphaerium naegelianum Unger	*Gomphosphaeria lacustris* Chodat
Merismopedia tenuissima Lemmerman	*Microcystis aeruginosa* Kuetz.
Nostoc commune Vauch.	*Oscillatoria* sp.
Rhabdoderma sp.	*Schizothrix calcicola* (Ag.) Gom.
Phylum: Chlorophyta	
Dictyosphaerium sp.	*Eudorina elegans* Ehrenberg
Pandorina sp.	*Pediastrum duplex* Meyern.
Scenedesmus quadricauda (Turp.)	*Spondylosium* sp.
Staurastrum sp.	
Phylum: Pyrrhophyta	
Ceratium hirudinella (O.F. Muell.)	

REFERENCES

Allen, T.F.H., Bartell, S.M. and Koonce, J.F., 1977. Multiple stable configurations in ordination of phytoplankton community change rates. Ecology, 58: 1076–1084.

APHA, 1985. Standard Methods for the Examination of Water and Wastewater (16th Edition). American Public Health Association, American Water Works Association, and Water Pollution Control Federation, Washington, DC, 1268 pp.

Bayne, C.K., Beauchamp, J.J., Begovich, C.L. and Kane, V.E., 1980. Monte Carlo comparisons of selected clustering procedures. Pattern Recognition, 12: 51–62.

Ehinger, W.J., 1988. Plankton composition and temporal variation among the three basins of Lake Whatcom, Washington. Master's thesis, Huxley College of Environmental Studies, Western Washington University, Bellingham, WA, 133 pp.

EPA, 1983. Methods of Chemical Analysis of Water and Wastes. EPA-600/4-79-020.

Fisher, D. and Langley, P., 1986. Conceptual clustering and its relation to numerical taxonomy. In: W.A. Gale (Editor), Artificial Intelligence and Statistics. Addison-Wesley, Reading, MA, pp. 77–116.

Gauch, H.G., Whittaker, R.H. and Wentworth, T.R., 1977. A comparative study of reciprocal averaging and other ordination techniques. J. Ecol., 65: 157–174.

Goodman, L.A. and Kruskal, W.H., 1954. Measures of association for cross classifications. J. Am. Stat. Assoc., 49: 732–745.

Kenkel, N.C. and Orloci, L., 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology, 67: 919–928.

Legendre, P., Dallot, S. and Legendre, L., 1985. Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. Am. Nat., 125: 257–288.

Lighthart, B., Kraft, G.F. and Flora, C.J., 1972. The limnology of Lake Whatcom. Tech. Rep. 15, Institute for Freshwater Studies, Western Washington State College, Bellingham, WA.

Lind, O.T., 1985. Handbook of Common Methods in Limnology. Kendall/Hunt, Dubuque, IA, 197 pp.

Lorenzen, C.J., 1967. Determination of chlorophyll and pheo-pigments: spectophotometric equations. Limnol. Oceanogr., 12: 343–346.

Ludwig, J.A. and Reynolds, J.F., 1988. Statistical Ecology. Wiley-Interscience, New York, 337 pp.

Matthews, G. and Hearne, J., 1991. Clustering without a metric. IEEE Trans. Pattern Anal. Machine Intell., 13 (forthcoming).

Pielou, E.C., 1984. The Interpretation of Ecological Data. A Primer on Classification and Ordination. Wiley-Interscience, New York, 263 pp.

Scavia, D. and Robertson, A., 1979. Perspectives on Lake Ecosystem Modeling. Ann Arbor Science, Ann Arbor, MI, 326 pp.

Reprinted from

Canadian Journal of Fisheries and Aquatic Sciences

Réimpression du

Journal canadien des sciences halieutiques et aquatiques

Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream

G. B. MATTHEWS, R. A. MATTHEWS, AND B. HACHMÖLLER

Volume 48 • Number 11 • 1991

Pages 2184–2190

Canada

Fisheries and Oceans Pêches et Océans

Printed in Canada by The Runge Press Limited
Imprimé au Canada par The Runge Press Limited

Mathematical Analysis of Temporal and Spatial Trends in the Benthic Macroinvertebrate Communities of a Small Stream

G. B. Matthews

Computer Science Department, Western Washington University, Bellingham, WA 98225, USA

R. A. Matthews

Huxley College of Environmental Studies, Western Washington University, Bellingham, WA 98225, USA

and B. Hachmöller

Barfußerstraße 42, 3550 Marburg, Germany

Matthews, G. B., R. A. Matthews, and B. Hachmöller. 1991. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. Can. J. Fish. Aquat. Sci. 48: 2184–2190.

Macroinvertebrates were collected at four sites in Padden Creek, a small second-order stream in Whatcom County, Washington, USA. Two upstream sites were characterized by high densities of sensitive taxa, predominantly mayflies, stoneflies, and caddisflies, and two downstream sites showed high densities of tolerant taxa, especially true flies, annelids, *Baetis* mayflies, and gastropods. Despite the small sample size, some statistical techniques proved useful. The first two components of correspondence analysis were used to confirm the existence of both seasonal and spatial trends in the benthic macroinvertebrate populations of the stream. Neither component alone, however, ordinated the samples with respect to these trends. Combinations of the first two components were required. A standard clustering technique, *k*-means clustering with squared Euclidean distance, further confirmed the seasonal trend. Nonmetric clustering, not widely used in the analysis of ecological data, was necessary to confirm the spatial trend. Nonmetric clustering was also able to identify a small number of "significant" taxa, i.e. taxa that reliably served as indicators of spatial position on the stream.

On a effectué un échantillonnage des macroinvertébrés à quatre sites du ruisseau Padden, un petit cours d'eau de second ordre situé dans le comté Whatcom de l'État de Washington (É-U). Des densités élevées de taxons sensibles étaient caractéristiques des deux sites d'amont, en particulier des éphémères, des perles et des phryganes, tandis que les deux sites d'aval abritaient des densités élevées de taxons tolérants, surtout des mouches, des annélides, des éphémères du genre *Baetis* et des gastéropodes. Malgré la faible taille des échantillons, certaines méthodes statistiques se sont révélées utiles. Ainsi, les deux premières composantes de l'analyse factorielle de correspondance ont permis de confirmer l'existence de tendances saisonnières et spatiales dans les populations de macroinvertébrés benthiques du cours d'eau. Toutefois, ni l'une ni l'autre de ces composantes n'a permis d'effectuer une ordination des échantillons en ce qui concerne ces tendances, ordination obtenue toutefois par la combinaison des deux premières composantes. L'agglomération de moyennes *k* couplée à la distance euclidienne au carré, une technique agglomérative normalisée, a permis d'étayer cette tendance saisonnière. L'agglomération non métrique, rarement utilisée dans l'analyse de données écologiques, a été nécessaire pour confirmer la tendance spatiale. Cette dernière analyse a aussi permis d'identifier un faible nombre de taxons "significatifs", c'est-à-dire des taxons qui ont servi d'indicateurs fiables de la position spatiale dans le cours d'eau.

Received September 9, 1990
Accepted May 2, 1991
(JA718)

Reçu le 9 septembre 1990
Accepté le 2 mai 1991

One of the fundamental principles of mathematical ecology is that changes in the statistical makeup of the biota are reflections of changes in the physical environment. The dominance of certain taxa at a particular site or between sites can serve as a quantifiable record of the strength and direction of environmental changes (Faith and Norris 1989). In the ecology of streams, there are often two dominant environmental changes, one associated with time and the other with location (Green 1974). The benthic community varies with the season, and also with its spatial position in the stream. Many benthic macroinvertebrates have habitat requirements that correspond to longitudinal gradients, upstream to downstream. For example, because they require highly oxygenated waters, many stoneflies are restricted to headwater streams, which are often less polluted and more turbulent than downstream reaches (Hynes 1970; McCafferty 1981). Many other stream characteristics can be viewed as changing along this longitudinal gradient, due to the unidirectional downstream flow. This view of streams as gradients has influenced many of the fundamental theories on how streams function, including organic matter processing, macroinvertebrate community trophic structure, in-stream primary productivity, and nutrient cycling (see Minshall 1988 and Fisher 1983 for general reviews). However, the complex distributions and patterns exhibited by macroinvertebrates

make statistical confirmation of such relationships difficult. The problem of identifying reliable taxonomic indicators of environmental changes is even more difficult.

In this paper we used ordination by correspondence analysis and clustering by two techniques, *k*-means clustering and nonmetric clustering, to obtain statistical confirmation of the benthic macroinvertebrate response to both the longitudinal and the seasonal trends. Correspondence analysis is well documented in the literature (e.g. Gauch et al. 1977; Kenkel and Orloci 1986; ter Braak 1986), but, while ordination has been used extensively for finding and confirming terrestrial vegetation gradients (e.g. Minchin 1987), it has been used much less frequently to examine gradients in stream data (e.g. Green 1974; Culp and Davies 1980; Sheldon and Haick 1981; Schaeffer and Perry 1986; Faith and Norris 1989). *K*-means clustering is also widely used in many fields (Jain and Dubes 1988). Nonmetric clustering, described in the Appendix, is a new technique and has not been widely applied to ecological data although we have found it useful in a variety of aplications (Matthews and Hearne 1991; Mathews et al. 1991). We found that the combination of these three analytical techniques provided an excellent approach to our data set. The spatial and temporal trends were both revealed by correspondence analysis. The temporal trend was confirmed by *k*-means clustering which successfully separated samples by date, and the spatial trend was similarly confirmed by the nonmetric clustering which successfully separated samples by site.

The data we used for our analyses were collected from Padden Creek, a small second-order stream located adjacent to the city of Bellingham in Whatcom County, Washington. Hachmöller (1989) and Hachmöller et al. (1990) found that the macroinvertebrate fauna in Padden Creek showed distinct upstream and downstream distribution patterns. These distribution patterns were thought to be related to differences in the riparian community, especially canopy cover, and the input of nonpoint-source runoff from residential and agricultural areas, which created a turbid, nutrient-enriched "lower reach" in the creek.

Methods

Macronvertebrate Sampling

Four sites were sampled in Padden Creek (Fig. 1). Site 1 was located approximately 1 km downstream from the Lake Padden outfall in a forested, relatively undisturbed area. Site 2 was located in a channelized reach that had a less diverse substrate than Site 1. Both Sites 1 and 2 were upstream from the confluence of Padden and Connelly Creeks. Connelly Creek is a nutrient-enriched tributary that drains agricultural and residential lands. Site 3 was located about 1.5 km downstream from Connelly Creek in a forested city park that was more disturbed than Site 1. Site 4 was located in a freshwater wetland close to the mouth of Padden Creek. Based on vegetation, water quality, and substrate sampling, Hachmöller et al. (1990) and Uhlig (1991) characterized the four sites as in Table 1.

The macroinvertebrate samples were collected monthly at each site from June through October 1988 using a Surber sampler (1-mm net mesh). Ten samples were collected at each site on each date. The invertebrates were keyed to the lowest practical taxon (genus in most cases) using the following references: Anderson (1976), Edmunds and Jensen (1976), Hatch (1953–65), Jewett (1959), Merritt and Cummins (1984), Pennak (1978), Ricker and Scudder (1975), Ross (1937), Stark and Gaufin (1976), and Stone et al. (1965). Macroinvertebrate densities for each taxon were calculated as the average number of individuals per square metre (n = 10 per site and date).

Statistical Tests

Throughout this section, a "sample" refers to the pooled macroinvertebrate densities at a unique site and date: there were 20 samples in this study (4 sites × 5 dates). Individual macroinvertebrate densities for each taxon are called "replicates." There were 10 replicates for each taxon (63 taxa) at each date and time (a maximum of 12 600 replicates, many of which had values of zero). Some statistical tests were performed on both the sample data averaged by replicate and the raw data, not averaged by replicate; however, only the results from the averaged sample tests are reported here. Generally, as might be expected, the raw data yielded similar results, but with larger variances.

We ordinated the samples using correspondence analysis. Correspondence analysis (also called reciprocal averaging) determines taxa scores and sample scores in an "uninformed" manner, i.e. without prior grouping of the samples. Thus, samples are ordinated independently of information regarding the actual site or date at which they were collected. For our purposes, it was important that the correspondence analysis procedure give several ordinations of the samples (first, second, third components, etc.), for we found that two components were necessary to reveal trends indicated by our subjective evaluations. The correspondence analysis procedure is similar to principal components and factor analysis, but has been shown to be superior to these methods in typical environmental data sets (Kenkel and Orloci 1986; Ludwig and Reynolds 1988).

The data were also clustered by the *k*-means algorithm using squared Euclidean distance, and nonmetric clustering. *K*-means clustering (Jain and Dubes 1988) views the samples as points in n-dimensional space, where n is the number of taxa. It seeks "clusters" of samples such that the distance between samples from the same cluster is generally less than the distance between samples from different clusters. The measure of distance between samples is called the metric. A clustering is optimal in the metric sense if it maximizes the difference between the average intracluster distance and the average intercluster distance. There are many measures of "distance" for samples, and the choice of a particular distance metric can have a radical effect on the resulting clusters. For our *k*-means clustering we used squared Euclidean distance. Nonmetric clustering, described in the Appendix, is a new procedure that does not use a distance metric to determine clusters (Matthews and Hearne 1991). Instead, a clustering is optimal in the nonmetric sense if it maximizes the association between clusters and a large number of taxa. Each taxon is also given a "score" by nonmetric clustering, which is a measure of how strongly that particular taxon is associated with the clustering. Both nonmetric and *k*-means clustering are uninformed procedures, like correspondence analysis, and do not require prior grouping of samples.

Correspondence analysis, metric clustering, and nonmetric clustering were also used in an effort to identify diagnostic taxa, i.e. a subset of the taxa that could be used as indicators of environmental conditions. Correspondence analysis not only ordinates the samples, but also ordinates the taxa, and thus "large" taxa scores might be taken to indicate taxa important

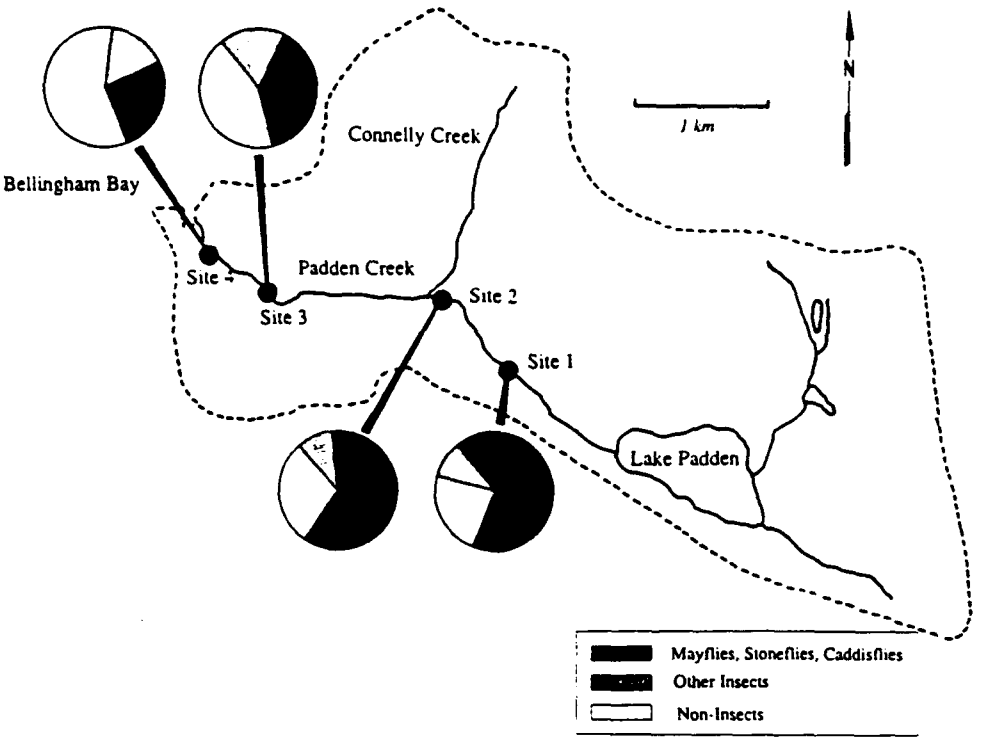


FIG. 1. Padden Creek sampling sites and relative proportions of major macroinvertebrate taxa. The macroinvertebrate proportions were calculated by averaging the macroinvertebrate densities (no./m^2) at each site for the entire study period.

TABLE 1. Characterization of the four sampling sites.

Factor	Site 1	Site 2	Site 3	Site 4
Nutrient concentration	Low–moderate	Low–moderate	Elevated	Elevated
Riparian vegetation	Second-growth coniferous forest	Alder; gaps in canopy	Second-growth coniferous forest	Freshwater wetland
Stream gradient	61 m/km	19 m/km	8 m/km	11 m/km
Substrate	Diverse cobble–pebble	Uniform cobble–pebble	Diverse pebble–sand	Diverse pebble–sand

to the correspondence analysis ordination. K-means clustering does not rank taxa in importance, and so was not used to identify diagnostic taxa. Nonmetric clustering, however, is designed to cluster data and simultaneously identify the taxa that are "important" with respect to these clusters (Matthews and Hearne 1991). In this regard it is similar to conceptual clustering techniques (Fisher and Langley 1986), which not only cluster the data, but attempt to show how those clusters can be characterized by a small subset of the data parameters. A nonmetric clustering which is meaningfully related to a spatial or longitudinal trend will also give a list of important taxa, which could be used as indicators of that trend.

Results

Hachmöller (1989) and Hachmöller et al. (1990) found that the most abrupt change in macroinvertebrate community structure occurred between Sites 2 and 3, which was attributed primarily to the influence of Connelly Creek. These changes can be seen in the pie charts summarizing the benthic community in Fig. 1. Mayflies, stoneflies, and caddisflies were collected in greater densities at the upstream sites (Sites 1 and 2); these three orders made up 62–67% of the macroinvertebrate densities at the upstream sites, but only 26–40% of the densities at the downstream sites (Sites 3 and 4). In addition, many of the uncommon taxa (less than 0.5% of the total density) were collected more frequently at the upstream sites, especially large, predatory stoneflies. This may be an artifact of the taxonomic technique because not all taxa were identified to the same level. In particular, Chironomidae and many of the noninsect taxa were identified only to family. This is a pervasive taxonomic dilemma, and its relevance to our statistical tests will be discussed below. In general, the macroinvertebrates collected at the downstream sites were mostly taxa having relatively cosmopolitan distributions such as *Baetis* and Chironomidae and included a large proportion of noninsect taxa such as oligochaetes, gastropods, etc.

Table 2 lists the average densities (number per square metre) for the most common taxa (greater than 0.5% of the total density) that were collected from Padden Creek from June through October 1988. A complete listing of the 63 Padden Creek taxa is given in Hachmöller (1989). It should be noted that the des-

TABLE 2. Macroinvertebrate densities and nonmetric clustering (NMC) scores for major taxa.

Padden Creek macroinvertebrate taxa	% total density	Average densities (no./m²)				NMC score
		Site 1	Site 2	Site 3	Site 4	
Plecoptera						
Malenka spp.	5.0	78.57	80.94	3.44	6.45	
Skwala spp.	0.6	3.87	11.84	1.72	2.15	0.47
Suwallia/Triznaka/Sweltsa complex	2.9	75.77	19.59	0.21	1.50	0.89
Ephemeroptera						
Baetis spp.	10.3	48.43	130.88	60.06	108.93	0.12
Cinygmula spp.	2.4	46.71	32.29	1.93	1.29	0.68
Epeorus spp.	3.9	95.58	33.58	1.29	0.21	
Ironodes spp.	2.3	36.38	32.50	4.73	3.01	0.47
Paraleptophlebia spp.	3.7	13.56	41.11	40.04	31.00	
Serratella spp.	1.9	2.36	57.04	4.52	1.50	
Trichoptera						
Glossosoma spp.	5.4	31.43	119.04	25.61	7.10	
Hydropsyche spp.	4.5	14.52	29.06	8.82	1.07	0.33
Rhyacophila spp.	1.1	24.54	10.33	1.29	0.64	0.89
Parapsyche spp.	4.8	8.18	12.70	107.63	33.15	−0.47
Diptera						
Chironomidae	7.7	60.70	87.83	59.20	54.03	0.26
Simuliidae	4.0	17.00	0.21	53.17	64.79	−0.33
Amphipoda						
Gammarus lacustris	0.7	0.00	0.86	5.59	17.43	−0.80
Annelida						
Enchytraeidae	31.4	210.97	260.05	237.88	355.42	−0.26
Lumbriculidae	2.2	3.44	9.25	23.03	39.61	−0.68
Gastropoda						
Ferissia	0.5	0.00	3.87	10.97	2.79	−0.41
Gyraulus	1.4	0.00	13.99	9.68	24.54	−0.26

ignation of "common" is somewhat arbitrary because, again, not all taxa were identified to the same level.

Confirmation of the observed longitudinal and seasonal trends by correspondence analysis can be seen in Fig. 2, which plots all samples by the first two components of correspondence analysis. Neither trend, however, corresponds well with a single component of correspondence analysis. Instead, the seasonal differences tend to spread along a "northwest–southeast" line, and the longitudinal trends spread along an orthogonal, north-east – southwest" line. We believe that this observation is important, as the emphasis in much statistical ecology is on recognizing a single, dominant gradient in the population. This is the motivation behind "detrended" correspondence analysis, for example, which attempts to force a one-dimensional ordination for data sets. In our case, a two-dimensional ordination was essential.

The ordinations by correspondence analysis led to difficulties in the identification of indicator taxa. First, as seen in Fig. 2, neither of the first two sample score components, alone, corresponds with the trends of interest. Each is a combination of both trends. Accordingly, neither of the first two taxa scores could be used to determine indicator taxa for either trend. Second, although correspondence analysis taxa scores were partially associated with the trends (for example, positive taxa scores were generally assigned to "upstream" taxa and negative taxa scores to "downstream" taxa) the correspondence analysis scores were strongly influenced by rare taxa. Only three of the top 20 correspondence analysis taxa scores were from common taxa, these three being *Hydropsyche*, *Malenka*, and *Serratella*.

The seasonal trend was confirmed by *k*-means clustering, which separated the samples by date. The June and July samples were placed in one cluster and the August, September, and October samples in the other, except for one August sample which was placed in with the June and July samples (see Fig. 2a). On the other hand, nonmetric clustering confirmed the observed longitudinal trend, and clustered all upstream (Sites 1 and 2) samples into one cluster and all downstream (Sites 3 and 4) samples into the other cluster (see Fig. 2b).

The attempt to identify indicator taxa using nonmetric clustering was very successful. Unlike correspondence analysis, most of the top taxa scores produced by nonmetric clustering were from common taxa. These 15 out of the 20 top scores were common taxa, and are listed in Table 2. This was impressive considering there were only 20 common taxa and that nonmetric clustering is "naive" in that it did not use total macroinvertebrate density as a selection criterion. Further, we verified the robustness of this taxonomic subset using a "leave-one-out" strategy. The nonmetric clustering taxa scores were recalculated based on only 19 "training" samples, leaving one sample out, and then the group (upstream or downstream) for the omitted sample was predicted using taxa scores generated from the other 19 samples. This procedure was repeated with each sample being the one omitted, obtaining 20 tests; thus we obtained an estimate of the rate at which errors might occur in using these taxa scores to classify unknown samples, by simply counting the number of the "left-out" samples that were mis-

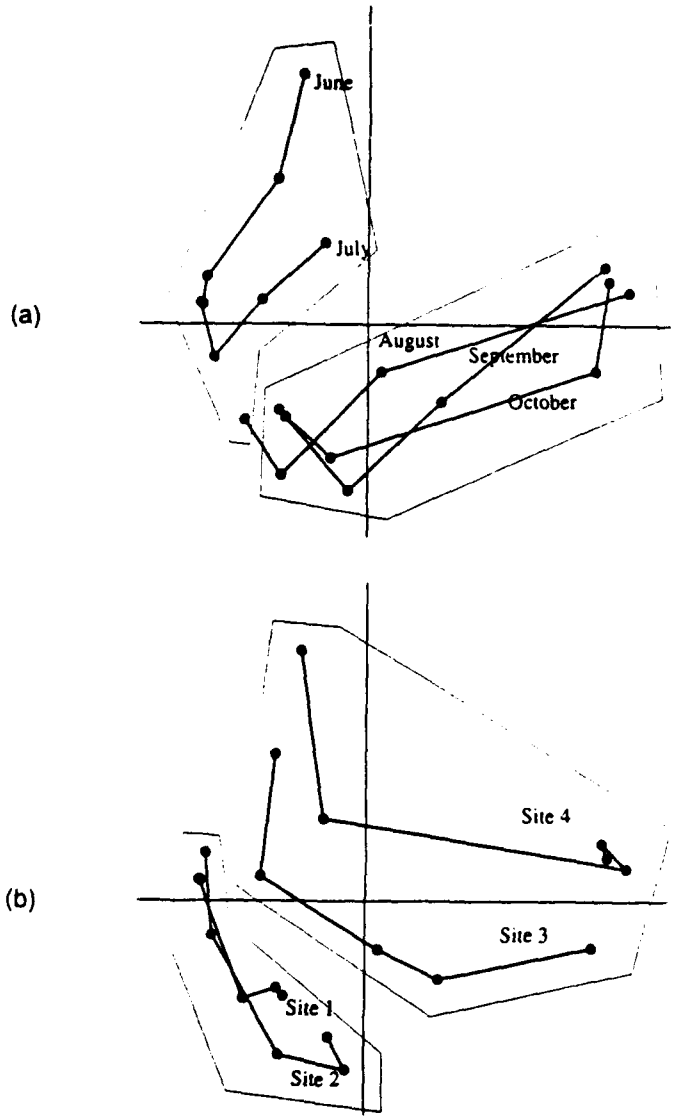


FIG. 2. Samples plotted with respect to the first two components of correspondence analysis. In Fig. 2a, heavy lines connect samples from a single date; in Fig. 2b, heavy lines connect samples from a single site. The "northwest–southeast" trend in dates and the "northeast–southwest" trend in sites are illustrated. Grouping of samples in Fig. 2a is by *k*-means clustering with squared Euclidean distance; grouping of samples in Fig. 2b is by nonmetric clustering.

classified. For our nonmetric clustering-derived characterization, there were no erroneous classifications. By comparison, we also performed "leave-one-out" testing using a linear discriminant procedure to reclassify the left-out sample. The linear discriminant misclassified 15% (3 out of 20), and this was in spite of the linear discriminant being an "informed" procedure, i.e. input to the linear discriminant procedure consisted of both the data points and an identification of which data points came from upstream samples and which from downstream samples. Nonmetric clustering is, in contrast, an "uninformed" procedure. Input to the nonmetric clustering procedure consisted only of the data points, and no information about the location of the samples. Nonmetric clustering was able to deduce the locations of the samples from the macroinvertebrate densities alone.

Discussion

Our statistical analyses supported our initial hypothesis that there were longitudinal and seasonal trends evident in the macroinvertebrate data. Ordination of samples by correspondence analysis was clearly possible (Fig. 2); however, a two-dimensional ordination was necessary to confirm each of the one-dimensional trends.

The existence of (at least) two gradients in a data set made interpretation of the data by clustering more difficult. Our two clustering techniques yielded radically different clusters because the structure of the data was complex enough to warrant two interpretations. Which trend is the "strongest" depends on how "strongest" is interpreted. In our professional judgement, the most obvious trend was the longitudinal trend. There were marked differences in the makeup of the macroinvertebrate communities from upstream and those from downstream. However, the existence of this "obvious" trend was not confirmed by *k*-means clustering. Instead, a rather new tool, nonmetric clustering, that approaches data clustering from radically different assumptions was required to "confirm the obvious."

The fact that correspondence analysis gave high scores to rare taxa might be expected because, if a taxon is rare, and only shows up at one site or date, it will, of course, be highly correlated with that site or data. But many factors can affect the reported densities of rare taxa, including drift and emergence as well as sampling technique, sorting, and taxonomic experience. Because only some of these factors are associated with a gradient, correspondence analysis may not be robust in data sets where there are many uncommon taxa. In Padden Creek, 43 of the 63 taxa were uncommon, i.e. making up less than 0.5% of the total density. The conclusion we draw is that taxa scores from correspondence analysis should not be viewed individually or in small subsets (such as the top 20), but only collectively.

Nonmetric clustering was the only technique that proved successful in both (a) confirming an observed trend and (b) providing a set of indicator taxa for that trend. Nonmetric clustering identified a subset of 15 common taxa, given in Table 2, that provided enough information to classify the samples, and did so more accurately than a linear discriminant.

Conclusion

Ecologically the dominant trends in our stream data were the longitudinal trend, where, typically, mayflies, stoneflies, and caddisflies were found at the upstream sites (Sites 1 and 2), while noninsects and tolerant taxa were found at the downstream sites (Sites 3 and 4), and the seasonal trend. Our subjective judgement was that the longitudinal trend was more significant in this study than the seasonal one. Correspondence analysis ordination of the macroinvertebrate data from Padden Creek confirmed the presence of both the longitudinal and seasonal trends in the taxa, but only as a "mixture" of each of the first two components of the ordination. In addition, correspondence analysis typically gave rare taxa the highest taxa scores, even though their relevance to large-scale trends in the data set was minor. *K*-means clustering favored the seasonal trend over the longitudinal trend, while nonmetric clustering favored the longitudinal trend. The nonmetric clustering also provided a robust means of simplifying the description of upstream and downstream clusters by identifying a set of 15 of the most com-

mon taxa that could be used to ordinate samples in other studies if a reduced sampling effort was desirable. This set of 15 proved to be a robust indicator of the location of the sample, regardless of the season in which the sample was collected. Nonmetric clustering has not previously been used to analyze benthic macroinvertebrate data, but should prove to be a useful tool for future studies, with broad applications and major advantages over current techniques.

References

ANDERSON, N. H. 1976. The distribution and biology of the Oregon Trichoptera. Technical Bulletin 134, Agricultural Experimental Station, Oregon State University, Corvallis, OR.

CULP, J. M., AND R. W. DAVIES. 1980. Reciprocal averaging and polar ordination as techniques for analyzing macroinvertebrate communities. Can. J. Fish. Aquat. Sci. 37: 1358–1365.

EDMUNDS, D. F., AND D. L. JENSEN. 1976. The mayflies of North and Central America. University of Minnesota Press, Minneapolis, MN.

FAITH, D. P., AND R. H. NORRIS. 1989. Correlation of environmental variables with patterns of distribution and abundance of command and rare freshwater macroinvertebrates. Biol. Conserv. 50: 77–98.

FISHER, D., AND P. LANGLEY. 1986. Conceptual clustering and its relation to numerical taxonomy, p. 77–116. *In* W. A. Gale [ed.] Artificial intelligence and statistics. Addison-Wesley, Reading, MA.

FISHER, S. G. 1983. Succession in streams, p. 7–27. *In* J. R. Barnes and G. W. Minshall [ed.] Stream ecology, application and testing of general ecological theory. Plenum Press, New York, NY.

GAUCH, H. G., R. H. WHITTAKER, AND T. R. WENTWORTH. 1977. A comparative study of reciprocal averaging and other ordination techniques. J. Ecol. 65: 157–174.

GREEN, R. H. 1974. Multivariate niche analysis with temporally varying environmental factors. Ecology 55: 73–83.

GOODMAN, L. A., AND W. H. KRUSKAL. 1954. Measures of association for cross classifications. Am. Stat. Assoc. J. 49: 732–764.

HACHMOLLER, B. 1989. The benthic macroinvertebrate community in Padden Creek, summer, 1988. M.S. thesis, Western Washington University, Bellingham, WA.

HACHMOLLER, B., R. MATTHEWS, AND D. BRAKKE. 1990. Effects of riparian community structure, sediment size, and water quality in a small, suburban stream. Northwest Sci. 65: 125–132.

HATCH, M. H. 1953–65. The beetles of the Pacific Northwest. Vol. I–IV. University of Washington Press, Seattle, WA.

HYNES, H. B. N. 1970. The ecology of running waters. University of Toronto Press, Toronto, Ont.

JAIN, A. K., AND R. C. DUBES. 1988. Algorithms for clustering data. Prentice Hall, Englewood Cliffs, NJ.

JEWETT, S. G. 1959. The stoneflies (Plecoptera) of the Pacific Northwest. Oregon State College Press, Corvallis, OR.

KENKEL, N. C., AND L. ORLOCI. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 67: 919–928.

LUDWIG, J. A., AND J. F. REYNOLDS. 1988. Statistical ecology. John Wiley, New York, NY.

MATTHEWS, G. B., AND J. W. HEARNE. 1991. Clustering without a metric. IEEE Trans. Pattern Anal. Mach. Intell. 13: 175–184.

MATTHEWS, R. A., G. B. MATTHEWS, AND W. J. EHINGER. 1991. Classification and ordination of limnological data: a comparison of analytical tools. Ecol. Model. 53: 167–187.

MCCAFFERTY, W. P. 1981. Aquatic entomology. Jones and Bartlett Publishers, Inc., Boston, MA.

MERRITT, R. W., AND K. W. CUMMINS. 1984. An introduction to the aquatic insects of North America. 2nd ed. Kendall/Hunt Publishing Company, Dubuque, IA.

MINCHIN, P. R. 1987. An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69: 89–107.

MINSHALL, G. W. 1988. Stream ecosystem theory: a global perspective. J. N. Am. Benthol. Soc. 7: 263–288.

PENNAK, R. W. 1978. Freshwater invertebrates of the United States. John Wiley and Sons, New York, NY.

RICKER, W. E., AND G. G. E. SCUDDER. 1975. An annotated checklist of the Plecoptera (Insecta) of British Columbia. Syesis 8: 333–348.

ROSS, H. H. 1937. Nearctic alder flies of the genus *Sialis*. Ill Nat. Hist. Surv. Bull. 21: 57–78.

SCHAEFFER, D. J., AND J. A. PERRY. 1986. Gradients in the distribution of riverine benthos. Freshwater Biol. 16: 745–757.

SHELDON, A. L., AND R. A. HAICK. 1981. Habitat selection and association of stream insects: a multivariate analysis. Freshwater Biol. 11: 395–403.

STARK, B. P., AND A. R. GAUFIN. 1976. The nearctic genera of Perlidae. Misc. Publ. Entomol. Soc. Am. 10(1).

STONE, A., C. W. SABROSKY, W. W. WIRTH, R. H. FOOTE, AND J. R. COULSON. 1965. A catalog of the Diptera of America north of Mexico. Agricultural Research Service, United States Department of Agriculture, Washington, D.C.

TER BRAAK, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67: 1167–1179.

UHLIG, L. J. 1991. Comparison of predator-prey relationships between stoneflies and mayflies in various habitats of Padden Creek. M.S. thesis, Western Washington University, Bellingham, WA.

Appendix: Nonmetric Clustering

We give here a brief introduction to the technique of nonmetric clustering, which is described fully in (Matthews and Hearne 1991). Traditional clustering algorithms, such as *k*-means clustering, rely on a metric, or distance measure, defined over *n*-dimensional space. Points are then divided into clusters based on cluster "quality," where quality is in turn based on simultaneously minimizing intracluster distance and maximizing intercluster distance. In Fig. A.1, for instance, the points in the upper right would constitute one cluster because they are all close to each other, and the points in the lower left would constitute the second cluster because they are all close to each other and at the same time far from the points in the other cluster.

Problems arise with this method when other dimensions are added, however. In Fig. A.2a, the points all have the same x and y coordinates as in the previous figure, but a random value for the z dimension has been added. Intuitively, the points are still in the same clusters, and the third dimension represents pure noise that should be ignored. Metric-based clustering, however, must compose a metric out of all dimensions, with the result that the clusters proposed for the data are as shown in Fig. A.2b. If metric-based clustering is to succeed at all, some kind of data transformations or weighted metrics must be employed.

Nonmetric clustering, on the other hand, is not based on an *n*-dimensional metric. Instead, each dimension is examined independently of the others, and the association between the clustering and the dimension is measured. In Fig. A.1 the association between the obvious clusters and each of the x and y axes is evident. A quantitative measurement of this association is used to indicate the strength of the association. Guttman's λ (Goodman and Kruskal 1954), which is similar to a chi-squared statistic, is used for reasons discussed by Matthews and Hearne (1991). The optimal clustering, then, is selected as the one that has the strongest association with the largest number of dimensions. The dimensions themselves are not combined into a metric, and there is no call to include all dimensions in the estimate of clustering quality.

For our example data set, the nonmetric clustering for three dimensions, shown in Fig. A.2c is identical to the obviously "correct" clustering in two dimensions. This is because the best associations between clustering and dimensions are with the x and y axes. There is no way a clustering can be found that will associate well with more than two axes, and so only the x and y axes are used to measure clustering quality, and the z axis is ignored.

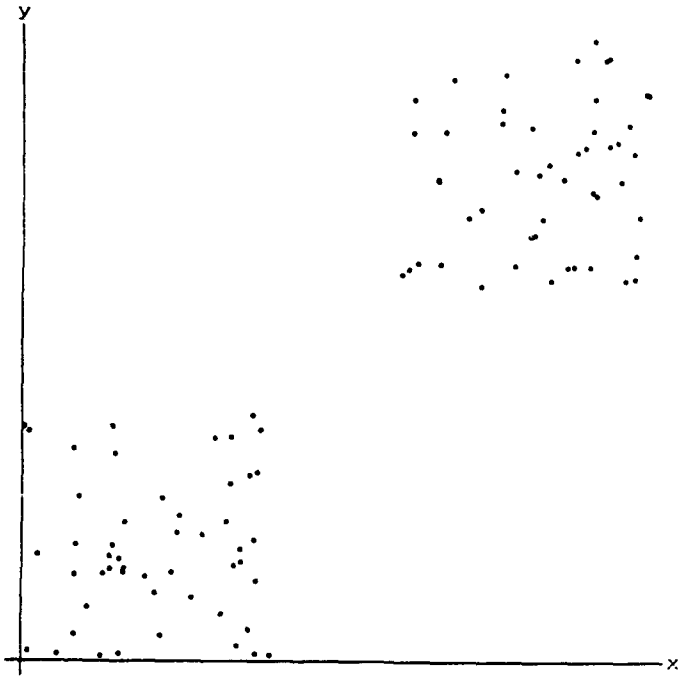


FIG. A.1. Artificially generated data set clustered in two dimensions.

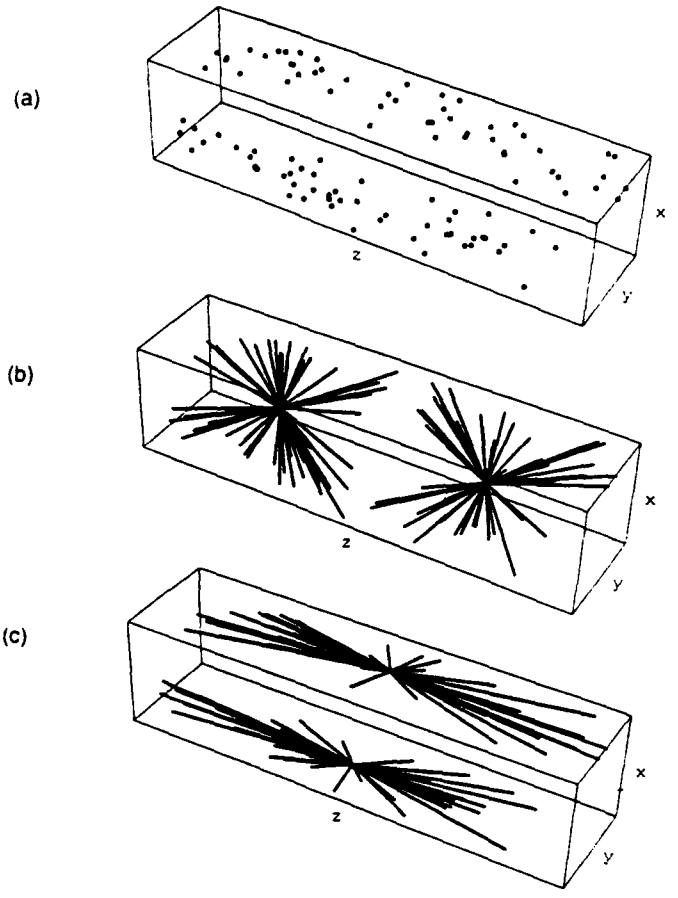


FIG. A.2. Artificial data set of Fig. A1 with (a) a random component in the z dimension added, (b) clustering by k-means, and (c) nonmetric clustering.

A computer program, called RIFFLE, implementing nonmetric clustering, has been constructed and is described in Matthews and Hearne (1991). RIFFLE was used for all nonmetric clustering discussed in this paper.

Nonmetric clustering thus offers the following advantages over traditional methods: (1) it does not combine counts from dissimilar taxa by means of sums of squares, or other ad hoc mathematical techniques; (2) it does not require transformations of the data, such as normalizing the variance; (3) it works without modification on incomplete data sets; (4) it can work without further assumptions on different data types (e.g. species counts or presence/absence data); (5) significance of a taxon to the analysis is not dependent on the absolute size of its count, so that taxa having a small total variance, such as rare taxa, can compete in importance with common taxa, and taxa with a large, random variance will not automatically be selected, to the exclusion of others; (6) it provides an integral measure of "how good" the clustering is, i.e. whether the data set differs from a random collection of points; and (7) it can, in some cases, identify a subset of the taxa that serve as reliable indicators of the physical environment; in our case, the indicator species were proved, in testing, to be more reliable than indicators based on a linear discriminant.

The primary disadvantages of nonmetric clustering, as we see them, are as follows. (1) There are some cases, documented in (Matthews and Hearne 1991), where metric clustering is to be preferred over nonmetric clustering. In general, we recommend using both, and examining the results critically, rather than accepting a single clustering method as the best for all cases. (2) The RIFFLE implementation of nonmetric clustering is very computer intensive, and takes much longer to run than k-means clustering. (3) Implementations of the technique, such as RIFFLE, are not widely available yet.

Environmental Sciences, 2, 3 (1993) 000-000
MYU, Tokyo

ES000

Multivariate Analyses of the Impact of the Turbine Fuel Jet-A Using a Standard Aquatic Microcosm Toxicity Test

W. G. Landis, R. A. Matthews, A. J. Markiewicz, N. J. Shough and G. B. Matthews*

Institute of Environmental Toxicology and Chemistry
Huxley College of Environmental Studies
Western Washington University
Bellingham, WA 98225 USA
*Computer Science Department
Western Washington University
Bellingham, WA 98225 USA

(Received September 9, 1992; accepted July 24, 1993)

Key words: multispecies toxicity testing, Jet-A, microcosm, nonmetric clustering, multivariate analysis

We investigated the toxicity of the water soluble fraction (WSF) of the turbine fuel Jet-A using the standard aquatic microcosm (SAM) method. The SAM experiment was conducted using concentrations of 0, 1, 5 and 15% WSF in 3 L SAMs containing species of organisms. The toxicant was added on day 7 of the 63-day experiment. Physical, chemical, and biological measurements were collected twice each week from day 11 through day 63. In the highest WSF treatment group an algal bloom ensued, generated by the toxicity of the WSF to *Daphnia*. As the test proceeded, the *Daphnia* populations increased and the algal populations decreased to about the reference values. In the last few weeks of the experiment *Cyprinotus* (ostracod) densities were higher in the reference than in the other treatment groups and *Philodina* (rotifer) densities were lower in the reference than in the other treatment groups. Because of high sampling variance, the ANOVA results suggested that few of these effects were significant. Multivariate analyses, however, revealed two distinct divergences between treatment groups: an early divergence that was probably due to the *Daphnia*/algae response, and a late divergence that was much more subtle, and may have been related to changes in the detrital quality in the different treatment groups. The variables that were most important in distinguishing the four treatments shifted during the course of the experiment, demonstrating the fallacy of

using only one index or a few measured endpoints in the evaluation of community-level interactions.

1. Introduction

Multispecies toxicity tests are usually referred to as microcosm or mesocosm tests, although a clear definition of these terms has not been put forth. Multispecies toxicity test systems range from approximately 1 L (e.g., mixed flask cultures) to thousands of liters, as in the case of the pond mesocosms used in pesticide registration testing. In the standardized aquatic microcosm (SAM) method[1] developed by Taub and colleagues,[2-12] the composition of the microcosm is clearly defined (Table 1). In other types of microcosms, the physical, chemical, and biological composi-

Table 1
Summary of test conditions for conducting the SAM Jet-A toxicity test.

Organisms:	Algae added on Day 0 at 10^3 cells for each taxa: *Anabaena cylindrica*, *Ankistrodesmus* sp., *Chlamydomonas reinhardi* 90, *Chlorella vulgaris*, *Lyngbya* sp., *Scenedesmus obliquus*, *Selenastrum capricornutum*, *Stigeoclonium* sp., and *Ulothrix* sp. Animals added on Day 4 at concentrations in parentheses: *Daphnia magna* (16), *Cypridopsis* sp (ostracod) (6), *Hypotricha* (protozoa) (0.1/ml), *Philodina* sp. (rotifer) 0.03/ml)
Test vessel:	One-gallon (3.8 L) glass jars; 16.0 cm wide at the shoulder; 25 cm tall witnh 10.6 cm openings
Medium:	T82MV; 3 L added to each container
Sediment:	Autoclaved silica sand (200 g), ground, crude chitin (0.5 g), and cellulose powder (0.5 g) added to each container
Replication:	6 replicate microcosms × 4 treatments
Reinoculation: (each microcosm)	Once per week add one drop (~0.05 ml) to each microcosm from a mix containing 5×10^2 cells of each alga
Addition of test materials:	Test material added on day 7 by removing 450 ml from each container and then adding appropriate amounts of the WSF to produce concentrations of 0, 1, 5 and 15 percent WSF. After toxicant addition the final volume was adjusted to 3 L
Test duration:	63 days
Temperature:	20° to 25°C
Light intensity:	80 μE m^2 photosynthetically active radiation/s (850 to 1000 fc)
Photoperiod:	12 h light/12 h dark
Sampling frequency:	2 times each week
Measurements:	Algal, invertebrate and protozoa counts, pH, dissolved oxygen, optical density. Calculated parameters included species concentrations, DO, DO gain and loss, net P/R ratio, pH, algal species diversity, *Daphnia* fecundity, algal biovolume, and biovolume of available algae

tion may vary widely.

Typically, the goals of multispecies toxicity tests are to detect changes in the population dynamics of the individual taxa that would not be apparent in single-species tests, and to detect community-level differences that are correlated with treatment groups. One of the major difficulties in the evaluation of multispecies toxicity tests has been to analyze the complex data set on a level consistent with these goals. A number of statistical approaches have been used to evaluate multispecies toxicity data. Analysis of variance (ANOVA) is the classic method used to examine differences between the treatment groups. However, because multispecies toxicity tests generally run for weeks, or even months, there are problems with using ANOVA, including the increased likelihood of a Type II error (accepting a false null-hypothesis), the presence of temporal dependence among the variables, and the difficulty of graphically representing the results. Conquest and Taub[13] developed a method to overcome some of the problems by using intervals of nonsignificant difference (INDs). This method corrects for the likelihood of Type II errors and produces intervals that are easily graphed. The method is routinely used to examine data from SAM toxicity tests, and is applicable to other multivariate toxicity tests. The major drawback is that this method can only be used to examine one variable at a time. While this addresses the first goal in multispecies toxicity testing, it ignores the second.

Multivariate data analysis methods are necessary to address the second goal of detecting community-level differences. One of the first multivariate methods used in toxicity testing was the calculation of ecosystem strain developed by Kersting[14–16] for a relatively simple (three species) microcosm. At about the same time, Johnson[17,18] developed a multivariate algorithm using the n-dimensional coordinates of a multivariate data set and the distances between these coordinates as a measure of divergence between treatment groups. Both of these methods have the advantage of examining the ecosystem as a whole rather than by single variables. A major disadvantage of both these multivariate methods (and of many others) is that all of the data are usually incorporated without regard to measurement units or the appropriateness of including all variables, even random ones, in the analysis.

Ideally, a multivariate statistical test used for evaluating complex data sets will have the following characteristics: (i) it will not combine counts from dissimilar taxa by means of sums of squares, or other *ad hoc* mathematical techniques; (ii) it will not require transformations of the data; (iii) it will work without modification on incomplete data sets; (iv) it will work without further assumptions on different data types (e.g., species counts or presence/absence data; (v) the significance of a taxon to the analysis will not depend on its abundance, so rare taxa can compete in importance with common taxa; (vi) it will provide an integral measure of "how good" the analysis is (i.e., whether the data set differs from a random collection of points); (vii) it will, in some cases, identify a subset of the taxa that serve as reliable indicators of the physical environment. To our knowledge, only one multivariate technique (nonmetric clustering) satisfies all these criteria.[19]

In this paper, we use ANOVA (with INDs) and three multivariate techniques to search for meaningful patterns in data from a SAM toxicity test using the water soluble fraction (WSF) of Jet-A turbine fuel. Jet-A is one of the most widely available aviation fuels, and, because of its stringent manufacturing specifications, is an excellent choice for evaluating the effects of a complex organic toxicant on a multispecies system. The multivariate techniques include two conventional tests based on the ratio of multivariate metric distances (Euclidean and cosine of the vector distances), and one relatively new procedure, nonmetric clustering and association analysis.[19] All three of the multivariate techniques have proven useful in analyzing complex ecological data sets.[20-22]

2. Materials and Methods

2.1 *Reagents*

All chemicals used in the culture of the organisms and in the formulation of the microcosm media were reagent grade or as specified in the ASTM protocol.[1] Glassware for the preparation of the WSF of Jet-A was washed in nonphosphate soap, rinsed, soaked in 2N HCl for at least 1 h rinsed ten times with distilled water, dried, and autoclaved for 30 min. Jet-A was provided by Fliteline Services of Bellingham, Washington, U.S.A., and refined by Chevron. The sample was obtained from the sample valve used for quality control and water sampling to prevent contamination by the refueling apparatus. The shipment lot was recorded and is on file. Microcosm medium T82MV was used for extracting the soluble fraction of Jet-A. Twenty-five ml of Jet-A were added to a 1 L separatory funnel containing 1000 ml of T82MV medium. For 1 h, the mixture was repeatedly shaken for 5 min and allowed to stand for 15 min. The mixture was then allowed to stand overnight. The following day all but the upper 100 ml of the T82MV/WSF mixture was drained into a clean, sterile 1 L amber glass bottle and capped with a Teflon-lined screw cap. The WSF was used within 24 h or stored at 4°C for no longer than 48 h.

2.2 *Gas chromatography of WSF*

A gas chromatographic analysis of the WSF was carried out using a Tekmar LSC 2000 purge and trap (P&T) concentrator system in tandem with a Hewlett-Packard 5890A gas chromatograph and a flame ionization detector (FID).[23-25] Instrument blanks and deionized, distilled water blanks were used to verify the cleanliness of P&T and GC columns prior to analysis of the WSF samples. A 5 ml sample was injected into a 5 ml sparger, purged with prepurified nitrogen gas for 11 min and dry purged for 4 min. Volatile hydrocarbons, purged from the sample and collected on the Tenax/Silica gel column, were desorbed at 180°C directly onto the SPB-5 fused silica capillary column (30 m × 0.53 mm, ID 1.5 μm film). The column was held at 35°C for 2 min, increased to 225°C at 12°C/min, and held at that temperature for 5 min. A Spectra-Physics 4290 integrator was used to record the FID signal output of the volatile hydrocarbons that were separated and eluted from the

column by molecular weight.

2.3 *Short-term toxicity tests*

In order to determine the appropriate WSF concentrations to be used for the SAM microcosm, a series of short-term toxicity tests were performed. These included 96 h algal growth inhibition tests using three species of algae (*Chlamydamonas reinhardii*, *Ankistrodesmus falcatus*, and *Selenastrum capricornutum*) and a 48 h *Daphnia magna* acute toxicity test.

The test algae were grown in a semi-flow through culture apparatus on the microcosm media T82MV and collected during log-phase growth for inoculation into the test flasks. Five hundred ml Erlenmeyer flasks were used as test chambers. Each test chamber contained 100 ml of the following treatments (reps = 2/treatment): 0 (reference), 6.25, 12.5, 25, 50 and 100% WSF. All dilutions of the WSF were made using T82MV. The test organisms were added at a concentration of approximately 3.0×10^4 cells/ml. Test mixtures were incubated at 20.0°C ± 1.0°C, with a 12:12 h light/dark cycle. Cell densities were determined every 24 h during the 96 h test period using a Neubauer counting chamber. The cell numbers were plotted against the WSF concentrations. If possible, a least-squares regression line was drawn and the IC_{50} (concentration resulting in 50% inhibition compared to the control) was determined. Significant differences between groups were determined using ANOVA.

Daphnia magna 48 h acute toxicity tests[26] were conducted using T82MV medium at concentrations of 0, 6.25, 12.5, 25, 50 and 100% WSF (reps = 2/treatment). Ten neonates were placed in 250 ml beakers containing 100 ml of test solution. After 24 and 48 h, the numbers of dead cells were recorded. Data were analyzed graphically and statistically to obtain an estimate of the EC_{50}.

2.4 *SAM toxicity test*

The 63-day SAM protocol[1] was modified to allow dosing with the WSF. The WSF was added on day 7 by stirring each microcosm, removing 450 ml from each container, and adding WSF to produce concentrations of 0, 1, 5, and 15% WSF. The final volume was readjusted to 3 L using T82MV. No attempt was made to filter and retain the organisms withdrawn during the removal of the 450 ml prior to addition of the toxicant. All graphs and statistical analyses began with the next sampling day (day 11). Table 1 summarizes the organisms, conditions and modifications used for the Jet-A experiment.

2.5 *Data analysis*

The variables that were measured or calculated included the numerical densities for each species, dissolved oxygen (DO), DO gain and loss, net photosynthesis/respiration ratio (P/R), pH, algal species diversity, algal biovolume, and biovolume of "available" algae (i.e., available for consumption by filter feeders).[1] The ANOVA INDs[13] and the average values for each variable were plotted by treat-

ment group against time to identify significant differences. In addition, three multivariate clustering and significance tests were used to determine dose/response relationships. Two of the clustering procedures were based on the ratio of metric distances (Euclidean and cosine of vectors) within treatment groups vs between treatment groups. The third test used nonmetric clustering and association analysis.[19]

The biotic parameters used for the multivariate analyses are listed in Table 2. Treating each sameple on a given day as a vector of values, $x = \langle x_1 \cdots x_n \rangle$, with one value for each of the measured biotic variables, allows Euclidean distance between two sample points x and y to be computed as:

$$\sqrt{\sum_i (x_i - y_i)^2}.$$

The cosine of the vector distance between x and y can be computed as:

$$1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum y_i^2}}.$$

Subtracting the cosine from one yields a distance measure, rather than a similarity measure, with the measure increasing as the points get farther from each other.

The statistical significance of the metric clustering results was calculated using

Table 2

Biotic parameters used in the multivariate statistical tests.

Anabaena
Ankistrodesmus
Chlamydomonas
Chlorella
Daphnia
Ephipia
Small *Daphnia*
Medium *Daphnia*
Large *Daphnia*
Hypotricha (Protozoa)
Lyngbya
Miscellaneous sp.
Cyprinotus (Ostracod)
Philodina (Rotifer)
Scenedesmus
Selanastrum
Stigeoclonium
Ulothrix

Derived variables (e.g., diversity) were not used because they are not independent.

the within-between (W/B) ratio and an approximate randomizatioh test.[29] For each date, one sample point x was obtained from each of six replicates in the four treatment groups, giving a 24 × 24 matrix of distances. After the distances were computed, the ratio of the average within group distance (W) to the average between group distance (B) was computed (W/B). If the points in a given treatment group were, on average, closer to each other than they were to points in a different treatment group, then this ratio will be small. The significance of the ratio was estimated using an approximate randomization test.[29] This test is based on the null hypothesis that assignment of points to treatment groups is random, the treatment having no effect. Accordingly, the test repeatedly (500 times) assigned the 24 points randomly to (pseudo) groups and calculated the W/B ratio. If the null hypothesis is false, the randomly derived W/B ratio will be larger, on average, than the W/B ratio obtained from the actual treatment groups An estimate of the probability under the null hypothesis was obtained as $(n + 1)/(500 + 1)$, where n was the number of times the random W/B ratio was less than or equal to the actual W/B ratio.

In the nonmetric clustering and association test, the data were first clustered independently of treatment group, using the computer program RIFFLE.[22] Because the clustering analysis is naive to treatment group, the clusters may, or may not correspond to treatment effects. Under the null hypothesis, there should be no association between the clustering and the treatment groups. To test this hypothesis, the association between clusters and treatment groups was measured in a 4 × 4 contingency table, each point in treatment group i and cluster j being counted as a point in frequency cell ij. Significance of the association in the table was then measured with Pearson's χ^2 test:[30]

$$\chi^2 = \sum_{ij} \frac{(N_{ij} - n_{ij})^2}{n_{ij}}$$

$$n_{ij} = \frac{N_{+j} N_{i+}}{N},$$

where N_{ij} is the actual cell count, n_{ij} is the expected cell frequency obtained from the row (N_{+j}) and column (N_{i+}) marginal totals; and N is the total cell count (i.e., 24). The significance (probability under the null hypothesis) for this value of χ^2 was computed using standard procedures.[31]

3. Results

3.1 *GC analysis*

The results from the GC analysis of the WSF are shown in Fig. 1. Immediately after the WSF was added to the SAMS, approximately 50–60 peaks were distinguishable in the highest treatment group (15% WSF). By the end of the experiment, virtually all of the peaks had disappeared from the water column, probably

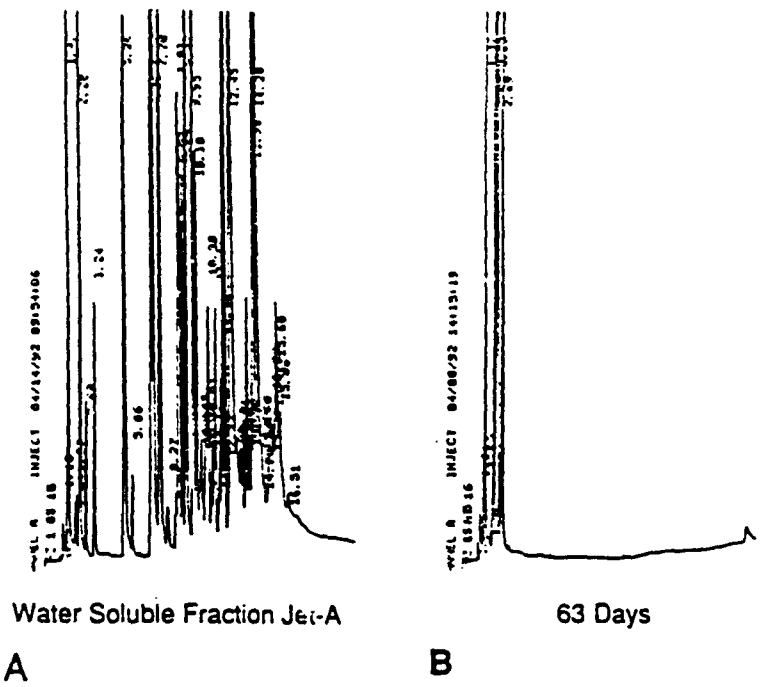


Fig. 1. Trap and purge GC chromatogram from the 15% WSF treatment group showing initial (Day 11) and final (Day 63) peaks.

due to volatization, photooxidation, biotransformation, and biodegradation.

3.2 *Short-term toxicity tests*

None of the 96 h acute algal toxicity tests indicated significant growth inhibition or enhancement correlated to treatment. However, the 48 h *D. magna* tests indicated that concentrations of 10–50% WSF caused *Daphnia* mortalities of 50–100%. The graphically derived EC_{50} was approximately 7% WSF (Fig. 2). Therefore, we expected that the highest concentration in the SAM experiments (15% WSF) would adversely impact the *Daphnia* populations shortly after the toxicant addition.

3.3 *SAM univariate results*

Daphnia population growth in the reference and lowest treatment group was similar throughout most of the experiment (Fig. 3). As expected, however, both of the higher treatment groups showed inhibition of *Daphnia* populations. In Treatment 3, the *Daphnia* populations (especially small *Daphnia*) started increasing on day 14. Treatment 4 did not show a major increase in the populations until day 17, and the population peak was not reached until after day 30.

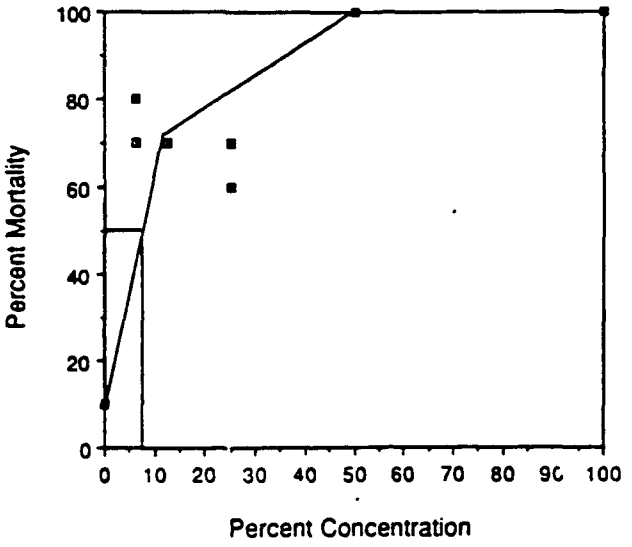


Fig. 2. 48-h acute *Daphnia* toxicity tests results for the WSF of Jet-A.

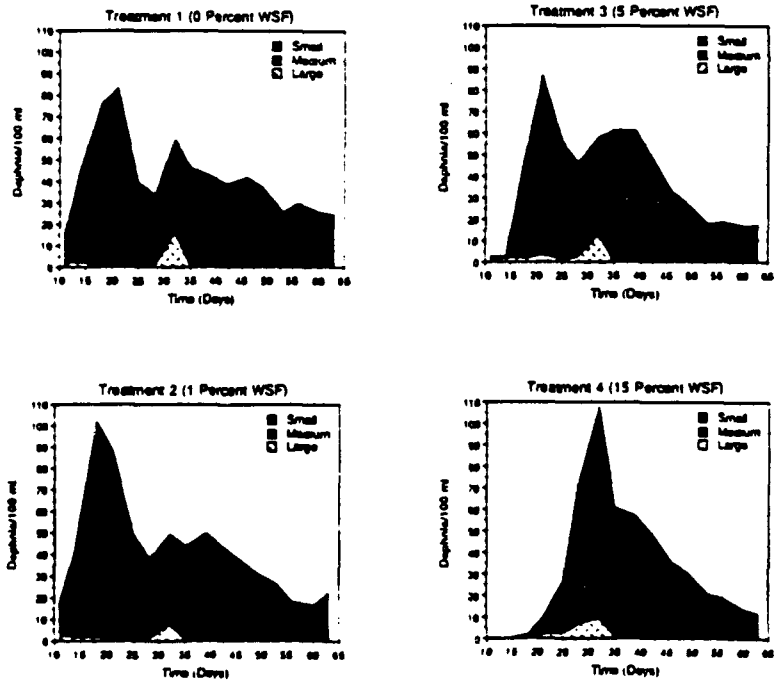


Fig. 3. *Daphnia magna* densities from the SAM toxicity test of the WSF of Jet-A.

Early algal blooms were observed in Treatments 3 and 4. (Fig. 4) ~~Figure 4 near~~ On day 21 the peak algal density in Treatment 4 was approximately four ~~here~~ times that of the reference. These increases were most likely due to reduced survival and reproduction in the *Daphnia* populations in the first few weeks of the experiment.

At the end of the experiment the average *Cyprinotus* (ostracod) density in the reference was approximately twice that of Treatment 4 (Fig. 5), and the population densities of other treatment groups were ranked in a dose/response manner. The ranking was consistent from day 49 onward. Because of the high sampling variance, the IND plots did not indicate any significant differences between treatments. Similarly, by the end of the experiment *Philodina* (rotifer), which were relatively uncommon throughout the experiment, were less numerous in the reference compared to Treatments 3 and 4. Again, because of the large sampling variance, the IND plots did not show any significant differences (Fig. 6).

The P/R ratio, measured by changes in daytime and nighttime DO concentrations, exhibited a dose: response relationship early in the experiment, with Treatments 3 and 4 being significantly different from the reference (Fig. 7a). The pH also responded in a dose/response manner to the addition of Jet-A. During the early part of the experiment (during the algal blooms), pH was significantly higher in

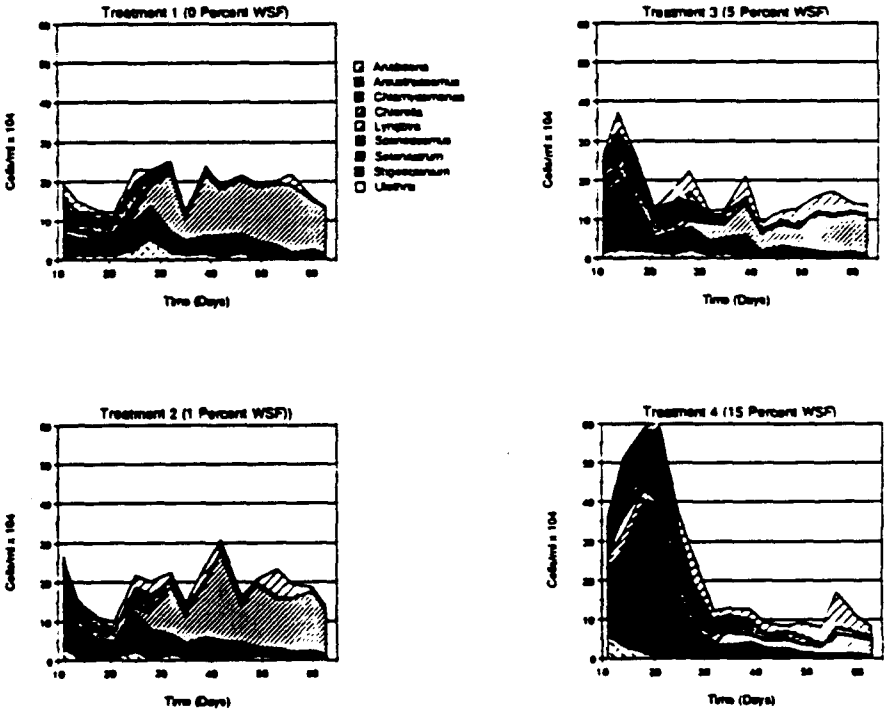


Fig. 4. Algal densities from the SAM toxicity test of the WSF of Jet-A.

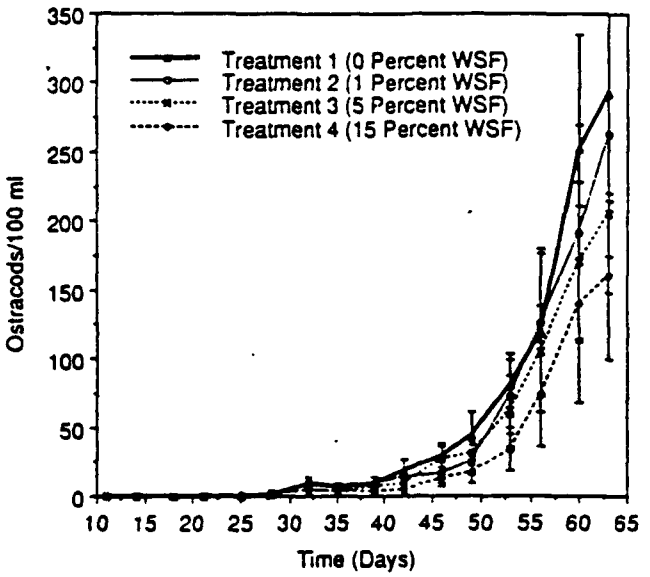


Fig. 5. *Cyprinotus* densities from the SAM toxicity test of the WSF of Jet-A.

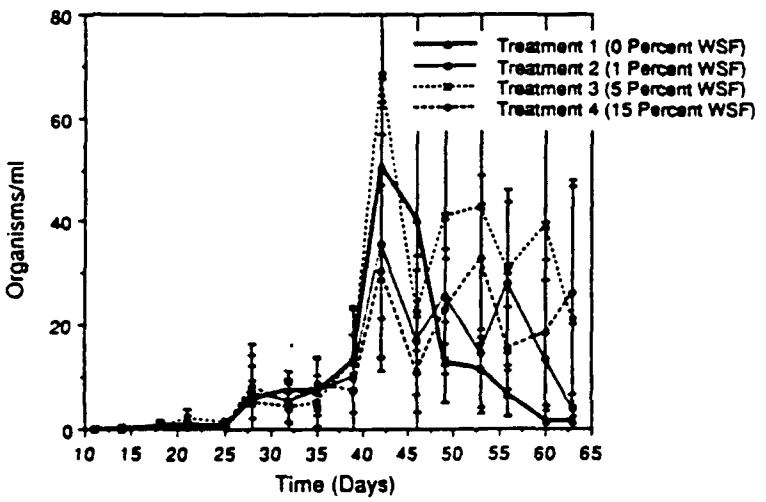


Fig. 6. *Philodina* densities from the SAM toxicity test of the WSF of Jet-A.

the two highest treatment groups than in the reference (Fig. 7b). On day 49 a second deviation from the reference was detected. No significant differences in pH were observed among the treatment groups by the end of the experiment.

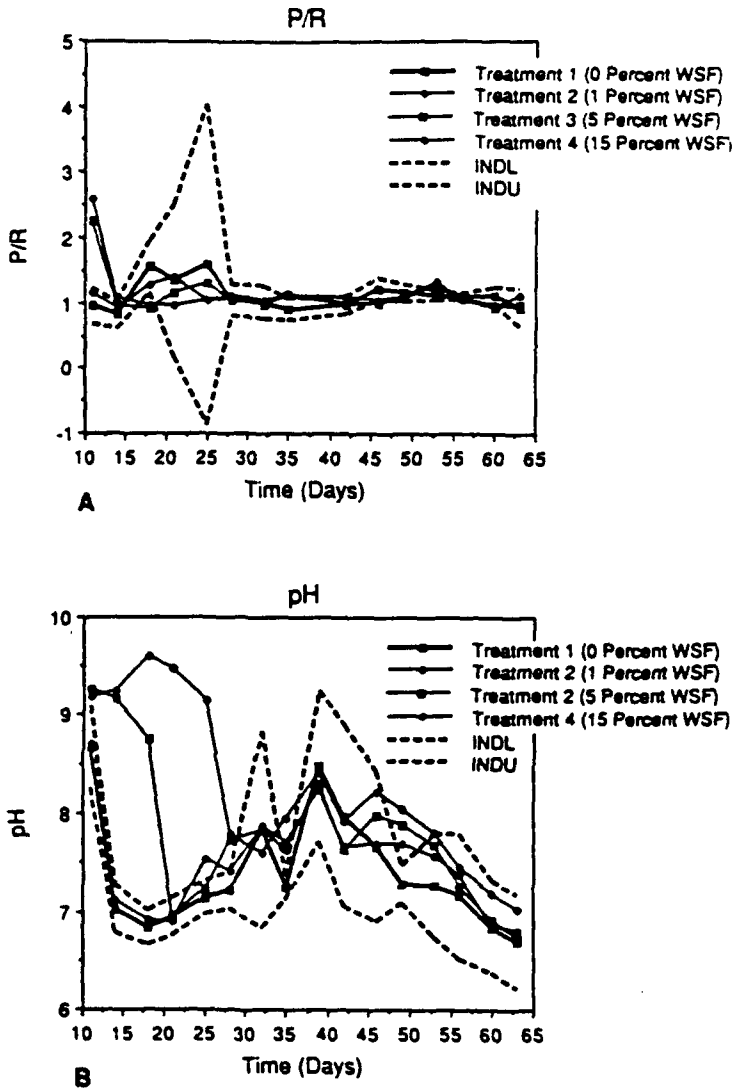


Fig. 7. Photosynthesis/respiration ratio and pH values from the SAM toxicity test of the WSF of Jet-A. Upper (INDU) and lower (INDL) limits of significance are shown as dashed lines.

3.4 *Multivariate results*

The significance levels for the three multivariate tests performed for each sampling day are graphed in Fig. 8. All three tests indicate that there were significant differences ($p \geq 0.95$) between treatment groups from day 11 through day 25, and again from day 46 through day 56. No consistent differences were observed from day 28 to day 39 and on days 60 and 63.

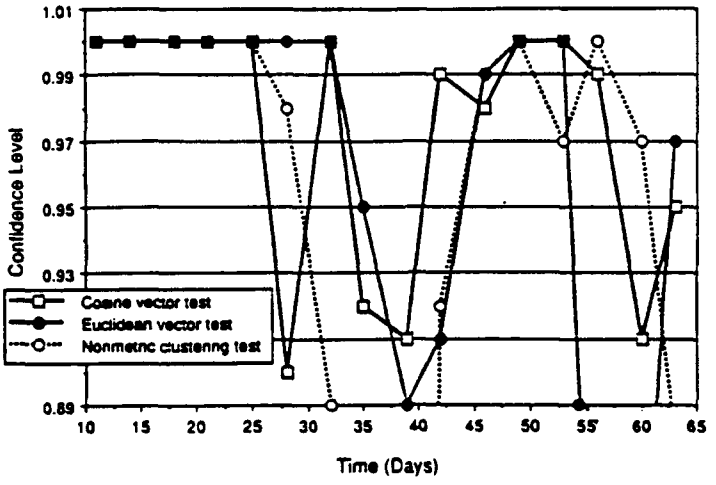


Fig. 8. Significance levels of three multivariate statistical tests (cosine vector, Euclidean vector, and nonmetric clustering) for the SAM toxicity test of the WSF of Jet-A.

In Fig. 9, the average cosine distances between the reference group and each of the three treatment groups are plotted on a log scale. The initial effect of the WSF dosing (day 11 to day 25) is apparent in the large distances between Treatment 1 and Treatment 4. Treatment 3 starts out distant from Treatment 1, but subsequently moves closer to the reference. The period of no significant differences (day 35 to day 46) is also obvious: none of the groups are especially far apart. During the second period of significant differences (day 46 to 56) a perfect dose/response relationship for all three treatments is seen, with higher doses becoming more distant from the control.

Using nonmetric clustering, we were able to list the variables that were the most important for separating the treatment group clusters for each day that measurements were collected (Table 3). This list shows that the specific variables that were most important for clustering changed over time. In addition, the number of variables used for clustering decreased from approximately 5–7 important variables on days 11–25 to ≤ 4 important variables from day 28 until the end of the experiment.

4. Discussion

Our examination of individual variables provided only a limited, and somewhat distorted view of the SAM response to Jet-A. The univariate data analysis did indeed show that there were some significant responses to the toxicant, especially during the first few weeks when the *Daphnia* populations declined and the algal

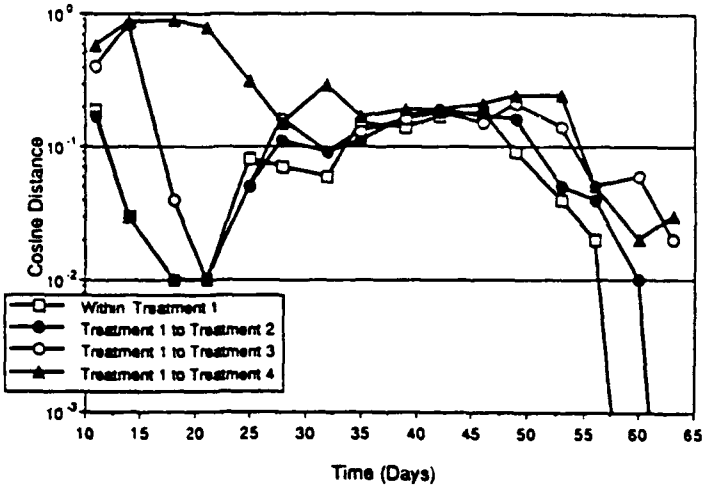


Fig. 9. Cosine distance from Treatment 1 to each of the remaining treatments for each sampling day. Smaller cosine distances indicate greater similarity between treatments.

populations peaked in the two highest treatment groups. However, the responses were scattered, and did not present a consistent pattern. Furthermore, the "significant" responses were actually gross aberrations of the microcosm, signifying wild

Table 3

Variables determined to be important in generating nonmetric clusters. Variables are listed in order of decreasing rank.

Day	*Important cluster variables (in rank order)*
11	M. *Daphnia, Chlorella, Chlamydamonas, Ulothrix*, S. *Daphnia, Selanastrum, Scenedesmus*
14	S. *Daphnia*, M. *Daphnia-Selenastrum*[1], *Chlamydamonas, Chlorella*, L. *Daphnia, Ankistrodesmus*
18	*Ankistrodesmus*, S. *Daphnia, Chlorella, Chlamydamonas, Selanstrum*, L. *Daphnia*
21	*Ankistrodesmus*, S. *Daphnia*, L. *Daphnia*-M. *Daphnia, Scenedesmus*
25	*Scenedesmus*, S. *Daphnia*, L. *Daphnia, Chlorella, Philodina*, M. *Daphnia*
28	*Ankistrodesmus*, L. *Daphnia, Scenedesmus*
32	S. *Daphnia*, M. *Daphnia, Ankistrodesmus, Chlorella*
35	*Ankistrodesmus*
39	M. *Daphnia-Selenastrum, Cyprinotus-Ankistrodesmus*
42	M. *Daphnia, Cyprinotus, Scenedesmus*
46	*Scenedesmus, Ankistrodesmus*, S. *Daphinia*, M. *Daphnia*
49	*Chlorella, Philodina, Ankistrodesmus, Lyngbya*
53	*Ankistrodesmus, Cyprinotus, Chlorella*
56	M. *Daphnia-Scenedesmus, Ankistrodesmus, Lyngbya*
60	*Lyngbya*, M. *Daphnia, Philodina, Chlorella*
63	*Chlorella, Ankistrodesmus, Philodina, Cyprinotus*

[1]Hyphen between variables denotes equal rank

swings in a taxon's population density. The confirmation of gross responses to a toxicant does not provide much more insight into the effects of the toxicant in an ecosystem than do short-term, single-species tests.

The multivariate statistics suggest a much more complex pattern of multiple divergences and convergences in the similarities between treatment groups. Much as an ecosystem could be expected to display the rise and fall of species assemblages, the SAMs appear to indicate that the first divergence was only the beginning of a series of responses.

The list of variables (Table 3) suggests that the first divergence, which occurred from about day 11 through day 32, resulted from predictable predator/prey interactions between *Daphnia* and algae. Theoretically, this divergence should be characterized by the following properties: (i) it should be fast, because the algae and *Daphnia* populations were introduced into the microcosm after being cultured in optimal laboratory conditions, in artificially high (and unstable) densities; (ii) it should be short-lived, because the populations are unstable in the nutrient-rich, early successional microcosm; (iii) there should be a tendency for the microcosms to drift away from their early treatment responses (especially because the WSF is essentially gone from the microcosms within a few days after its introduction) into more complex communities based on interactions between the remaining biotic constituents. This first divergence is the only type of response that is normally searched for in microcosm tests using conventional statistics, and is the response typically reported in SAM experiments.[9,10,32,33]

The second divergence occurred from about day 46 through day 60. During this time, other secondary consumers (e.g., *Cyprinotus* and *Philodina*) joined *Daphnia* and various algal taxa as being important in cluster development (see Table 3). The second divergence, therefore, may represent the long-term effects of the initial toxicant on a successionally more mature community. If so, the second divergence will be strongly influenced by detritus quality. Detritus is conditioned by bacteria and fungi, which are highly sensitive to toxins, but are not measured in the microcosm. Detritus that has passed through the gut of a consumer (e.g., *Daphnia*) is different from detritus that originates directly from unconsumed, dead algae. Therefore, the quality of the detritus may be highly affected by the treatment, but none of the factors influencing it are measured directly. Secondary consumers of detritus and bacteria (e.g., rotifers and ostracods) are no less affected by the quality of their food source than algal consumers, so the treatment-related alterations of the quality of detritus and bacteria will cause differences in the secondary consumer populations. Because this effect would occur late in the microcosm experiment and would be difficult to detect using univariate statistics, it would be easy to misinterpret as noise or as the effects of a degradation product.

Multiple divergences may also be explained without invoking direct impact of unseen biotic components of the system. The hypervolume defined by the multivariate data set for each treatment group may simply be moving in various directions and pass through the hypervolume of another treatment group at an ins-

tant in time. When viewed during that time, the two groups would appear similar (or to have "recovered"). In reality, this similarity is only a momentary confluence.

Taken separately, none of the biotic variables measured in the SAM experiment could clearly identify the second divergence. Even pH, a variable with a low sampling error, did not consistently distinguish the second divergence. Without corroboration, the few pH values that fell outside the INDs late in the experiment would probably have been considered outliers. However, the three multivariate analyses demonstrated a clear, significant dose/response relationship for both the first and second divergences. Nonmetric clustering was also able to select the variables that were important in distinguishing the four treatment groups, although the variables contributing to the differentiation changed from sampling day to sampling day (Table 3). These data suggest that reliance upon any one variable (e.g., *Daphnia*, or an index of variables, probably would have missed the second divergence. The implications are important. Currently, only small sections of the ecosystems are monitored and a heavy reliance is placed upon so-called indicator species. Our data suggest that such a practice could produce misleading interpretations because the best indicator species will most likely change over the course of an experiment a season, or site, etc.

In summary, we found at least two divergences between the similarities of treatment groups for the WSF of Jet-A. Multivariate analyses were crucial in identifying these patterns; conventional univariate statistics provided only clues. Furthermore, the complexity of the multivariate responses showed that reliance upon any particular set of indicator species may be misleading in determining the effects of stressors upon biological communities.

Acknowledgements

This research is supported by USAFOSR Grant No. AFOSR-91-2091 DEF. Robert V. Sahakian and Lester W. Keel participated in the preparation and sampling of the microcosms. Mark V. Haley, U.S. Army Chemical Research Division and Engineering Center kindly supplied the original cultures used in these experiments.

References

1 ASTM E1366-91. (1991): Standard practice for standardized aquatic microcosm: fresh water, volume 11.04. American Society for Testing and Materials, Philadelphia, PA.

2 Taub, F. B. (1969): Genotobiotic models of freshwater communities. *Verh. Internat. Verein. Limnol.* 17: 485–496.

3 Taub, F. B. (1976): Demonstration of pollution effects in aquatic microcosms. *Intern. J. Environ. Stud.* 10: 23–33.

4 Taub, F. B. and Crow, M. E. (1978): Loss of a critical species in a model (laboratory) ecosystem. *Verh. Internat. Verein. Limnol.* 20: 1270–1276.

5 Crow, M. E. and Taub, F. B. (1979): Designing a microcosm bioassay to detect ecosystem level effects. *Intern. J. Environ. Stud.* 13: 141–147.

6 Taub, F. B., Crow, M. E. and Hartmann, H. J. (1980): Responses of aquatic microcosms to acute mortality. In J. P. Giesy, Jr, Ed.: *Microcosms in Ecological Research*, Technical Information Center, U. S. Department of Energy, Washington, D.C., pp. 513–535.

7 Kindig, A. C., Loveday, L. C. and Taub., F. B. (1983): Differential sensitivity of new versus mature synthetic microcosms to streptomycin sulfate treatment. In W. E. Bishop *et al.*, Eds.: *Aquatic Toxicology and Hazard Assessment*: Sixth Symposium, ASTM 802, American Society for Testing and Materials, Philadelphia, PA, pp. 192–203.

8 Taub, F. B. and Read, P. L. (1983): Standardized Aquatic Microcosm Protocol: Final Report-Contract No. 223-80-2352. Vol II. Food and Drug Administration, Washington, D.C.

9 Taub, F. B., Kinding, A. C. and Conquest, L. L. (1987): Interlaboratory testing of a standardized aquatic microcosm. In: W. J.. Adams *et al.*, Eds.: *Aquatic Toxicology and Hazard Assessment*: 10th Volume, ASTM STP 971, American Society for Testing and Materials, Philadelphia, PA, pp. 385–405.

10 Taub, F. B., Kindig, A. C., Conquest, L. L. and Meador, J. P. (1988): Results of the interlaboratory testing of the standardized aquatic microcosm protocol. In G. Suter and M. Lewis, Eds.: *Aquatic Toxicology and Hazard Assessment*: Eleventh Symposium ASTM. American Society for Testing and Materials, Philadelphia, PA,

11 Taub, F. B. (1988): Standardized aquatic microcosm-development and testing. In A. Boudou and F Ribeyre, Eds.: *Aquatic Ecotoxicology*, Vol. II, CRC Press, Boca Raton, FL, pp. 47–92.

12 Taub, F. B. (1989): Standardized aquatic microcosms. *Environm. Sci. Technol.* 23(9): 1064–1066.

13 Conquest, L. L. and Taub, F. B. (1989): Repeatability and reproducibility of the standard aquatic microcosm: Statistical properties. In U. M. Cowgill and L. R. Williams, Eds.: *Aquatic Toxicology and Hazard Assessment*: 12th Volume, ASTM STP 1027, American Society for Testing and Materials, Philadelphia, PA, pp. 159–177.

14 Kersting, K. (1984): Development and use of an aquatic micro-ecosystem as a test system for toxic substances. Properties of an aquatic micro-ecosystem IV. *Int. Revue ges. Hydrobiol.* 69(4): 567–607.

15 Kersting, K. (1985): Properties of an aquatic micro-ecosystem. V. Ten years of observations of the prototype. *Verh. Internat. Verein. Limnol.* 22: 3040–3045.

16 Kersting, K. (1988): Normalized ecosystem strain in micro-ecosystems using different sets of state variables. *Verh. Internat. Verin. Limnol.* 23: 1641–1646.

17 Johnson, A. R. (1988): Diagnostic variables as predictors of ecological risk. *Environ. Manag.* 12(4): 515–523.

18 Johnson, A. R. (1988): Evaluating ecosystem response to toxicant stress: a state space approach. In W. J. Adams *et al.* Eds.: *Aquatic Toxicology and Hazard Assessment*: 10th Volume, ASTM STP 971, American Society for Testing and Materials, Philadelphia, PA, pp. 275–285.

19 Matthews, G. B. and Matthews, R. A. (1991): A model for describing community change. In: Pesticides in Natural Systems: How Can Their Effects Be Monitored? *Proceedings of the Conference, Environmental Research Laboratory*/ORD, Corvalis, OR. EPA 9109/9-91011.

20 Matthews, G. B., Matthews, R. A. and Hachmöller, B. (1991): Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Can. J. Fish: Aquat. Sci.* 48(11): 2184–2190.

21 Matthews, R. A., Matthns, G. B. and Ehinger, W. J. (1991): Classification and ordination of limnological data: a comparison of analytical tools. *Ecol. Model.* 53: 167–187.

22 Matthews, G. B. and Hearne, J. W. (1991): Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(2): 175–184.

23 ASTM D3710. (1986): goiling range distribution of gasoline and gasoline fractions by gas chromatography. *Annual Book of ASTM Standards* 5.03: 99–113.

24 ASTM D2887. (1986): Boiling range distribution of petroleum fractions by gas chromatography. *Annual Book of ASTM Standards*, 5.02: 658–667.

25 ASTM D-2 Proposal (P167). (1989): Manual on hydrocarbon analysis 4th Ed. Series MNL3. American Society for Testing and Materials, Philadelphia, PA.

26 ASTM E729-88a. (1988): Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphibians, volume 11.04 American Society for Testing and Materials, Philadelphia, PA.

27 Good, I. J. (1982): An index of separateness of clusters and a permutation test for its significance. *J. Statist. Comp. Simul.* 15: 81-84.

28 Smith, E. P., pontaseh, K. W. and Cairns, J., Jr. (1990): Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.* 24(4): 507-514.

29 Noreen, E. W. (1989): Computer Intensive Methods for Testing Hypotheses. Wiley-Interscience, New York, NY.

30 Fienberg, S. E. (1985): The Analysis of Cross-Classified Categorical Data. MIT Press, Cambridge, MA.

31 Press, W. H., Flannery, B. P., Teukolsky, A. A. and Vetterline, W. T. (1990): *Numerical Recipes in C, the Art of Scientific Computing*. Cambridge University Press, New York, NY.

32 Landis, W. G., Chester, N. A., Haley, M. V., Johnson, D. W., Muse, W. T., Jr. and Tauber, R. M. (1989): The utility of the standard aquatic microcosm as a standard method for ecotoxicological evaluation. In G. Suter and M. Adams, Eds.: *Aquatic Toxicology and Environmental Fate*: Eleventh Volume ASTM STP-1007, American Society for Testing and Materials, Philadelphia, PA, pp. 353-367.

33 Landis, W. R., Haley, M. V. and Chester, N. A. (In press): The use of the standardized aquatic microcosm in the evaluation of degradative bacteria in reducing impacts to aquatic ecosystems. In W.G. Landis, J. Hughes and M. Lewis Eds.: Environmental Toxicology and Risk Assessment: First Volume ASTM STP-1179. American Society for Testing and Materials, Philadelphia, PA.

Appendix 5

Clustering Without a Metric

T-PAM/13/2//40947

Clustering Without a Metric

Geoffrey Matthews
James Hearne

Reprinted from
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE
Vol. 13, No. 2, February 1991

Clustering Without a Metric

Geoffrey Matthews and James Hearne

Abstract—**We describe a methodology for clustering data in which a distance metric or similarity function is not used. Instead, clusterings are optimized based on their intended function: the accurate prediction of properties of the data. The resulting clustering methodology is applicable, without further ad hoc assumptions or transformations of the data 1) when features are heterogeneous (both discrete and continuous) and not combinable, 2) where some data points have missing feature values, and 3) where some features are irrelevant, i.e., have large variance but little correlation with other features. Further, it provides an integral measure of the quality of the resulting clustering. We have implemented a clustering program, RIFFLE, in line with this approach, and experiments with synthetic and real data show that the clustering is, in many respects, superior to traditional methods.**

Index Terms—**Clustering, cluster validity, multivariate data, proximity indexes, unsupervised learning.**

I. Introduction

THE goal of data analysis is the discovery of a model which fits the data. Statistical tools to accomplish this goal can differ in two ways: First, analysis tools differ in the kind of model which they fit to the data. For example, regression attempts to fit a linear subspace to the data points. Ordination attempts to fit a linear order to the data points. Clustering attempts to fit the data with a finite number of *clusters*, or subpopulations, each with distinct properties. We call this choice of model for an analytic tool its *model bias*. Second, analysis tools differ in the criteria used for goodness of fit. Regression typically seeks to minimize the sum of the squared distances of the data points from the regression subspace, but other measures, such as absolute value or a weighted sum, can be used. In clustering, the fitness criterion is usually the minimization of intracluster distance and simultaneous maximization of inter-cluster distance. The bias of the clustering procedure is then dependent on the distance function or metric used. We call this feature of an analysis tool its *fitness bias*.

We propose here a clustering methodology with a novel fitness bias. Our approach makes the clustering procedure easier to interpret and also leads to improved performance in some domains. Our rationale for the fitness bias is our concern for the uses of exploratory data analysis, and not an *a priori* judgement about similarity measures for data points. We assume that scientific data analysis is concerned with the patterns of cause and effect implicit in the data, and an appropriate analysis tool ought to be biased towards this in its model. In particular, a clustering methodology will attempt to find subpopulations of the data such that the observed data are highly contingent on the subpopulations. An optimal model of the data will be one which maximizes the *predictability* of data values, conditioned on the subpopulations. Our methodology thus maximizes the *utility* of the clustering, i.e., it attempts to minimize errors in predictions about samples from the data set. Further, we believe that it is particularly important in exploratory data analysis situations to *fit* the data without *distorting* the data, and our methodology therefore eschews all preprocessing of the data by, for example, normalization, substitutions for missing point-values, or elimination of outliers.

We take the distance metrics and functions, used in traditional clustering, to be ad hoc solutions to the problem of fitness bias, and inappropriate to most real world data analysis situations. Because we do not use distance functions, many of the problems of metric-based clustering do not arise. For instance, real scientific data sets are often heterogenous, or mixed, in their types. Some features of a data point may be categorical, others binary, and others real valued. To create a distance metric for such feature spaces introduces more ad hoc assumptions, or, worse, transforms the data to fit the analysis procedure. Secondly, incomplete data, i.e., data in which some or all points have *missing* values for some features, is common in real data sets. To use a distance metric on incomplete data requires some assumptions about the missing values, such as substitution of the mean, which again is a gross distortion of the original data. Thirdly, metric-based clustering cannot distinguish between important features, and those features in the data set which are noisy but which have no connection with the underlying cause and effect that determines the bulk of the other feature's values. Such "nuisance" features typically have to be filtered from the data set in advance of the clustering process. Finally, a measure of clustering *quality* is often not used, or is used separately from the clustering procedure itself. A clustering quality measure indicates not just which model fits "best," but provides some guidance on "how good" the fit actually is. This is critical, for instance, in deciding whether the data is better fit by two clusters, or by three clusters. Our methodology, however, incorporates a single measure of the *utility* of a clustering which 1) is meaningfully definable for continuous and discrete, ordered and unordered, feature types, 2) automatically ignores missing values in the data set, 3) automatically filters nuisance variables out of the eventual clustering, and 4) provides an integral measure of the quality of the clustering. Further, our approach is nonmetric (or, in statistical terms, nonparametric) in that our measures rely on the ordering of numeric data, but not the numeric distances.

We use our measure of utility, called *nonmetric fitness* (NMF) and described below in Section II-C, to guide a heuristic search over partitions of the data, seeking a global maximum. This approach is similar to conceptual clustering approaches to pattern recognition [1], in that the proposed usefulness of the clustering is an important factor in its fitness. Our approach is also similar Bayesian clustering [2], because we try to maximize the predictability of the actual data values, given the model. However, the system in [2] makes metric assumptions, that we do not, in assuming that the underlying distributions are multivariate Gaussian. Many of the tree-classifier systems [3]–[5] use fitness measures similar to NMF, but in classifier systems (using super-

Manuscript received August 15, 1989; revised August 23, 1990. This work was supported in part by the City of Bellingham, WA, and the Institute for Watershed Studies, Western Washington University.

The authors are with the Department of Computer Science, Western Washington University, Bellingham, WA 98225.

IEEE Log Number 9040947.

vised learning) instead of a clustering system (using unsupervised learning) such as ours.

Formal background for the approach is described below in Section II. Details of an implementation in the program, RIFFLE, are described in Section III. A comparison of its performance to that of the k-means clustering algorithm is described in Section IV-A, and some of the results of using the program on real world data are summarized in Section IV-B.

II. Formal Treatment

A. Clusterings

We assume the data constitute a set of I *points*, $D = \{x_i : i = 1, \cdots, I\}$ each of which is an ordered K-tuple, where K is the number of *features*, $x_i = \langle x_{i1}, \cdots, x_{iK} \rangle$. The features themselves will be named $P^1, \cdots, P^K$. The data can thus be viewed as a collection of I points in a K-dimensional space, the *feature space*. Each of the K features can be continuous or discrete, and may or may not have further structure, such as a natural zero or a natural unit (as in count data). Further, for each feature, "missing" or "unknown" can always be the value of a point.

A *clustering* of a data set D is a partition C of D into some number J of subsets, the *clusters* $C_1, \cdots, C_J$. The C_j are mutually exclusive and jointly exhaustive of D. Each data point $x_i \in D$ is given a number $j \in \{1, \cdots, J\}$ which is the number of the cluster to which x_i is assigned, and has no ordinal significance. We arbitrarily designate this feature, *cluster-number*, as the zeroth feature, so that the cluster-number for a data point x_i will be written x_{i0} and $x_{i0} = j$ iff $x_i \in C_j$. P^0 will then be another name for cluster-number.

B. Proportional Reduction in Error

We take the goal of a clustering to be accurate prediction of feature values for data points. We view cluster-number as simply another feature, and so we seek a quantitative measure of how well one or more features aid in the prediction of another. This is given by an estimate of the *reduction in error* achieved when using knowledge of the features, as opposed to prediction in ignorance. The measure we use is a generalization to an arbitrary number of features of Guttman's λ for two-dimensional cross-classification tables, which is extensively discussed in the literature [6]–[10]. The measure itself is only applicable to discrete features; our extension of it to clustering continuous features will be described in Section II-B-2.

1) Discrete Features: Consider the case where we are attempting to predict the value of one feature, P^3, on the basis of knowledge of two others, P^1 and P^2, and suppose that each of these features has three possible values, P^1 takes on values P_1^1, P_2^1, and P_3^1, and similarly for P^2 and P^3. With an adequate data set, we can obtain accurate frequency counts $f_{P_i^1 \wedge P_j^2 \wedge P_k^3}$ of the number of times a sample obtains values P_i^1, P_j^2, and P_k^3, for each i, j, and k. In other words,

$$f_{P_i^1 \wedge P_j^2 \wedge P_k^3} = |\{x : x_1 = P_i^1,\ x_2 = P_j^2,\ x_3 = P_k^3\}|.$$

See Fig. 1 where, for example, $f_{P_3^1 \wedge P_1^2 \wedge P_2^3} = 2$.

Now suppose for a particular data point x, we know $x_1 = P_3^1$ and $x_2 = P_1^2$, and wish to predict x_3. Clearly, we can do no better than look at all the frequency counts, for samples with the same values on P^1 and P^2, and choose the value of P^3 with the highest frequency. In other words, choose k such that $f_{P_3^1 \wedge P_1^2 \wedge P_k^3}$ is a maximum, which we denote: $\max_k \left(f_{P_3^1 \wedge P_1^2 \wedge P_k^3} \right)$. In Fig. 1, we have $f_{P_3^1 \wedge P_1^2 \wedge P_1^3} = 0$, $f_{P_3^1 \wedge P_1^2 \wedge P_2^3} = 2$, and $f_{P_3^1 \wedge P_1^2 \wedge P_3^3} = 5$, and so we should predict $x_3 = P_3^3$, and expect to be right about 5 out of 7 times.

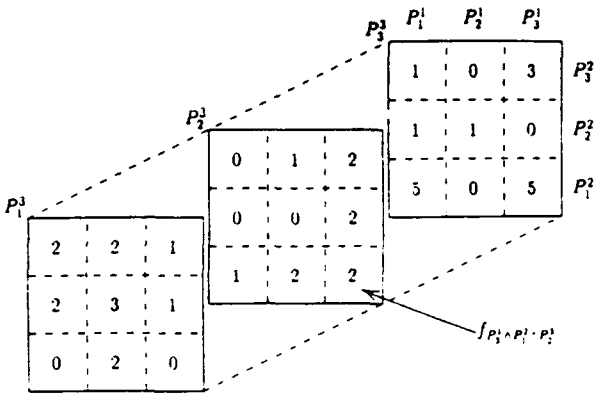


Fig. 1. A hypothetical frequency matrix for three features, P^1, P^2, and P^3, each with three possible discrete values. The frequency counts are entered in each cell, and the label for a typical cell illustrated.

If we make predictions for an entire collection of points, then our expected total correct percentage, in predicting P^3 on the basis of P^1 and P^2, would be

$$\text{Correct}(P^3 | \{P^1, P^2\}) = \frac{\Sigma_{i,j} \max_k \left(f_{P_i^1 \wedge P_j^2 \wedge P_k^3} \right)}{N}$$

where N is the total number of samples.

Generalizing to an arbitrary number of dimensions, an attempt to predict P^k, with values $P_{k'}^k$, conditioned on knowledge of a set of other features $\{P^i : i \in S\}$, $S \subseteq \{1 \cdots k-1, k+1, \cdots, K\}$, with values $P_{i'}^i$, is estimated to be correct with probability

$$\text{Correct}(P^k | \{P^i : i \in S\}) = \frac{\Sigma_{i'} \max_{k'} \left(f_{(\bigwedge_{i \in S} P_{i'}^i) \wedge P_{k'}^k} \right)}{N}.$$

If, on the other hand, we attempt to predict the value of a sample on a feature P^k using no information at all about the values of other features, then we can do no better than use the most common value of P^k, i.e., $P_{k'}^k$ where $\Sigma_{i'} f_{(\bigwedge_{i \in S} P_{i'}^i) \wedge P_{k'}^k}$ is a maximum, which we denote: $\max_{k'} \left(\Sigma_{i'} f_{(\bigwedge_{i \in S} P_{i'}^i) \wedge P_{k'}^k} \right)$. (The sums involved here are just the marginal totals of the frequency matrix.) If we use this for a guess, then our estimated probability of a correct prediction will be

$$\text{Correct}(P^k) = \frac{\max_{k'} \left(\Sigma_{i'} f_{(\bigwedge_{i \in S} P_{i'}^i) \wedge P_{k'}^k} \right)}{N}.$$

To obtain a measure of improvement based on these estimated probabilities, we can use the extent to which conditioning our predictions reduces error. The expected error rate in an unconditioned prediction is

$$\text{Error}(P^k) = 1 - \text{Correct}(P^k)$$

and the expected error rate in a conditional prediction is

$$\text{Error}(P^k | \{P^i : i \in S\}) = 1 - \text{Correct}(P^k | \{P^i : i \in S\}).$$

and the *proportional reduction in error* (PRE) is

$$\text{PRE}(P^k|\{P^i : i \in S\}) = \frac{\text{Error}(P^k) - \text{Error}(P^k|\{P^i : i \in S\})}{\text{Error}(P^k)}$$

$$= \frac{\Sigma_{i'}\left(\max_{k'}(f_{(\bigwedge_{i\in S} P^i_{i'})\wedge P^k_{k'}})\right) - \max_{k'}\left(\Sigma_{i'} f_{(\bigwedge_{i\in S} P^i_{i'})\wedge P^k_{k'}}\right)}{N - \max_{k'}\left(\Sigma_{i'} f_{(\bigwedge_{i\in S} P^i_{i'})\wedge P^k_{k'}}\right)}.$$

As a concrete example, we can calculate this quantity for the frequency matrix of Fig. 1, with the predicted feature, $k = 3$, and the known features $S = \{1,2\}$ as follows:

$$N = 39$$
$$\sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_1} = 13$$
$$\sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_2} = 10$$
$$\sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_3} = 16$$
$$\max_k \sum_{i,j} f_{P^1_i \wedge P^2_j \wedge P^3_k} = 16$$
$$\text{Correct}(P^3) = 16/39 \approx 41\%$$
$$\max_k f_{P^1_1 \wedge P^2_1 \wedge P^3_k} = 5$$
$$\max_k f_{P^1_1 \wedge P^2_2 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_1 \wedge P^2_3 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_2 \wedge P^2_1 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_2 \wedge P^2_2 \wedge P^3_k} = 3$$
$$\max_k f_{P^1_2 \wedge P^2_3 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_3 \wedge P^2_1 \wedge P^3_k} = 5$$
$$\max_k f_{P^1_3 \wedge P^2_2 \wedge P^3_k} = 2$$
$$\max_k f_{P^1_3 \wedge P^2_3 \wedge P^3_k} = 3$$
$$\sum_{i,j} \max_k f_{P^1_i \wedge P^2_j \wedge P^3_k} = 26$$
$$\text{Correct}(P^3|\{P^1, P^2\}) = 26/39 \approx 67\%$$
$$\text{PRE}(P^3|\{P^1, P^2\}) = \frac{26 - 16}{39 - 16} \approx 43\%.$$

In other words, the prediction of P^3 in ignorance will be correct 41% of the time, the prediction of P^3 using P^1 and P^2 will be correct 67% of the time, and so we can expect to be wrong 43% less often when we use P^1 and P^2 in the prediction of P^3 (assuming our sample is representative of the population).

If the set $\{P^i : i \in S\}$ contains only a single feature, then we write $\text{PRE}(P^k|P^i)$ for $\text{PRE}(P^k|\{P^i : i \in S\})$, and $\text{PRE}(P^k|P^i) = \lambda_{P^i P^k}$, Guttman's λ. Some properties of λ [6] are

1) λ lies between 0 and 1, inclusive, except when the entire population lies in a single cell of the table, in which case it is indeterminate.
2) λ is 1 if and only if all the population is in cells no two of which are in the same row or column.
3) If k and k' are independent, then λ is 0, but not necessarily vice versa.
4) λ is unchanged by permutations of rows or columns.

2) Continuous Features: To measure PRE on continuous features, discrete values are calculated from the continuous ones. The range of the feature is subdivided into J connected regions, and the discrete value of the continuous feature is the number of the region it falls into. This is justified by the observation that a clustering procedure, as a consequence of its model bias, produces only a finite number J of clusters. Even with a perfect clustering, feature predictions will be coarse, limited, for each feature, to one predicted value for each of the J clusters. On our model, we assume that each cluster will, accordingly, be associated with a single, connected subrange of each feature. For each J-clustering and for each continuous feature k, we choose $J - 1$ split-values, $s_{k_1} < \cdots < s_{k_{J-1}}$, and then define the discrete value for each sample x_i as

$$\text{discrete}_k(x_i) = j \text{ iff } s_{k_{j-1}} \le x_{ik} < s_{k_j}$$

where, for completeness, we can take $s_{k_0} = \min_i(x_{ik})$ and $s_{k_J} = \max_i(x_{ik}) + 1$. For example, with two clusters there will be a single split value and each data point will have either a "high" or a "low" value for each continuous feature. With three clusters, there would be "high," "medium," and "low" values. (More complex subsets could be imagined, but would greatly increase the complexity of the algorithm and, we believe, would find little use in practice.)

In any computation of PRE involving a continuous feature, it is understood that PRE is the maximum, over all such sets of split values, of the proportional reduction in error calculated in the usual way. Calculating such a maximum may involve a search over all candidate split values, or split values can be selected heuristically (as in our implementation, Section III), and these used as an approximation to the optimal split values.

C. A Nonmetric Measure of Clustering Fitness

The measure of error reduction PRE, defined above, can be used to define a measure of clustering fitness. The goodness of fit of a clustering is determined by how well feature values can be predicted using the clustering. Suppose, for example, we have a sample x, not part of the original data set, and we know only its feature values on the features in a given set, $\{P^i : i \in S\}$, and we want to predict a feature value x_j, with $j \notin S$. Using a given clustering of the data in this prediction is a two phase process. First, the cluster-number for x, i.e., x_0, is guessed, using the known feature values, and then the value of x_j is guessed, using x_0. The fitness of a clustering, therefore, can be measured by calculating $\text{PRE}(P^0|\{P^i : i \in S\})$ and $\text{PRE}(P^j|P^0)$. (We can, of

course, calculate $\text{PRE}(P^j|\{P^i : i \in S\})$ directly, but that answers a different question, regarding the intercorrelations of the features with each other. Here we seek an evaluation of the fitness of a clustering.) In a given clustering problem, we do not generally know j and S in advance, i.e., we do not know which features will be used in the prediction task. In fact, we take it to be part of the clustering task to determine which features *can* be used successfully in prediction. A data set, in other words, may be well clustered in some features, but also contain spurious or noisy features which have little relation to the clusters, and which could never be predicted accurately.

This leads to the following definitions. The *nonmetric fitness* (NMF^S) of a clustering C *in relation to* a feature set $\{P^i : i \in S\}$, is the average value of all terms of the form $\text{PRE}(P^0|\{P^i : i \in S'\})$ and of the form $\text{PRE}(P^j|P^0)$, where $S' \subseteq S$ and $j \in S$. A *particular feature set* (it need not be unique), for which NMF^S is a maximum is called an *optimal feature set* for C, and its nonmetric fitness is denoted simply by NMF. The fitness bias of our clustering methodology is toward clusterings with maximum NMF.

The introduction of the set S into our definition of clustering fitness, and the sets $S' \subseteq S$, permits further refinements in the notion of clustering fitness. Let the cardinality of S be $|S|$. If we restrict $|S|$ to be strictly less than the total number of features, $|S| < K$, we will obtain a clustering evaluated on a subset of features. Our fitness bias will then not only seek fit clusters, but will seek the best features for those clusters, resulting in "data reduction" on both the points (by grouping them into clusters) and on the features, but filtering out all but $|S|$ of them. On the other hand, if we restrict the size of S' in the definition of NMF we can control the amount of interdependence between features used to define the clustering. Setting $|S'| = 1$, for instance, requires the clustering to fit each feature in S independently of the others. Setting $|S'| = 2$ allows two-feature interactions, but excludes possible higher-order dependencies among features from consideration. (Both of these restrictions are provided as user options in our implementation, Section III.) The size of the optimal feature set $|S|$ is called the *number of significant features*, and the size of interactions allowed, $|S'|$, is called the *interaction-level*.

To illustrate the measure of clustering fitness, and as well the concomitant selection of split values to maximize NMF, consider the two-dimensional data of Fig. 2(a). We seek an optimal clustering into two clusters, with $|S| = 2$ (all features are significant), and $|S'| = 1$ (the interaction-level is one and we attempt to cluster on features independently). In Fig. 2(b) an optimal clustering and the two split values (dashed lines) are shown. The split values allow us to view the continuous features as discrete; each point will have either a "high" or "low" value on each feature, and consequently belong uniquely to one of the four cells of the frequency matrix. Points labeled 1 and 2 are clustered perfectly, because their value on any one feature P^0 (cluster-number), P^1 or P^2, determines the values on the other two. The point labeled "X", however, is more difficult. If it is assigned to cluster 1, then

$$\text{PRE}(P^0|P^2) = \text{PRE}(P^2|P^0) = 1.0$$

but

$$\text{PRE}(P^0|P^1) = \frac{(10+6)-11}{17-11} = 5/6$$

$$\text{PRE}(P^1|P^0) = \frac{(10+6)-10}{17-10} = 6/7.$$

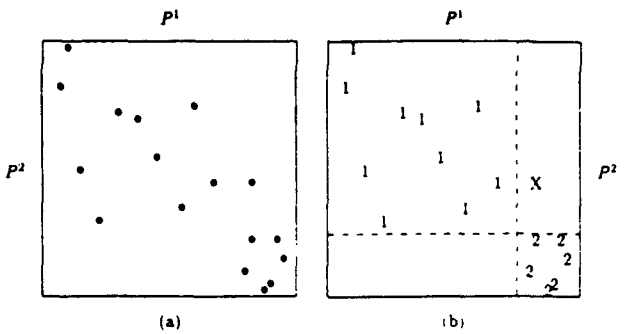


Fig. 2. An example data set (a) to be clustered using nonmetric fitness. Optimal clustering and split values are shown in (b). The point labeled "X" cannot be successfully clustered and will be assigned arbitrarily to cluster 1 or cluster 2.

On the other hand, if the point labeled "X" is assigned to cluster 2, then

$$\text{PRE}(P^0|P^1) = \text{PRE}(P^1|P^0) = 1.0$$

but

$$\text{PRE}(P^0|P^2) = \frac{(10+6)-10}{17-10} = 6/7$$

$$\text{PRE}(P^2|P^0) = \frac{(10+6)-11}{17-11} = 5/6.$$

In both cases, then, the NMF value will be $(1 + 1 + 6/7 + (5/6))/4 \approx 0.89$. The point labeled "X", therefore, can be assigned arbitrarily to either cluster, and both of the resulting clusterings are optimal. Any attempt to overcome this problem with the "X" point by adjusting the split values will create more problems than it solves, because more of the other points will then fall into one of the "troublesome" quadrants. This example also illustrates how maximization of PRE simultaneously on several different features, by adjusting their split values as well as the clustering, is necessary to achieve good fitness. Clustering one-dimensional, continuous-valued data on our criterion is a degenerate case, as any split value at all will give an NMF of 1.0 when cluster-numbers are selected to match discrete feature-values, and so we require $|S| \geq 2$.

III. Implementation

We have implemented our methodology in a computer program called RIFFLE, which is best described as a series of nested searches. The outermost loop searches for the best number of clusters ($J \geq 2$) simply by finding the best clustering for each number (in a user-specified range), and comparing the NMF values for each. One of the advantages of using NMF evaluations of clusterings is that fitness measures for clusterings with different J can be meaningfully compared. NMF is a measure of prediction accuracy, and whether one is predicting two values (high versus low) or three values (high, medium, or low), counts of correct and incorrect guesses can be compared (see Section IV-A-6, below).

The next level of search, given a fixed number of clusters J, is for the best cluster-numbering, i.e., assignment of points to clusters. Since a clustering is a partition of the points, the number of possible clusterings is $S(I, J)$ (Stirling numbers of the second kind, [11, pp. 90–91]), which prohibits exhaustive search. Instead, we begin with a random assignment of each

point to one of the J clusters, and then execute a hill-climbing search for improvements.

Currently this is done by reassigning a single point to a different cluster, recalculating NMF, and comparing the new fitness with the old. If any improvement is found, the point is left with its new cluster-number, otherwise the point is given its old cluster-number. In either case, other points are then examined to look for further improvement. Any time a point is successfully reassigned, *all* other points are then reexamined for possible further reassignment. This process continues until no improvements can be found by single-point reassignments, indicating we have reached a local maximum in NMF values. To avoid local maxima, the search may be repeated a number of times starting from a different initial random clustering. The number of repetitions necessary is, of course, domain dependent, but in practice we have never found more than about 50 to be necessary.

Nested within the search for optimal cluster-numbers is the evaluation of NMF, which involves a search for the optimal feature set and optimal split values for any continuous features in that set. User input relevant to this is the optimal feature set size $|S| = K^O$ and the interaction level $|S'| = K'$. If the interaction level is one, then for each feature P^k, we evaluate all terms of the form PRE($P^k|P^0$) and PRE($P^0|P^k$) and average these to give a "score" for P^k. If the interaction level is greater than one, all terms of the form PRE($P^k|P^0$) and PRE($P^0|S'$) are computed, for all sets S' with $|S'| = K'$. Each feature P^k is then given a score by averaging all terms in which it appears, either as the predicted feature or as a member of the set S'. In either case, those K^O features with the highest scores are selected to form the optimal feature set, which in turn is used to compute NMF. (For $K' > 1$ this procedure is heuristic, and optimality is not guaranteed.)

Finally, nested within the search for optimal features and calculation of NMF, is the search for optimal split values for the continuous features. Although there are infinitely many sets of split values, there are only finitely many that make a difference to a given data set. If $J-1$ split values are sought for a total of I points, $J-1$ distinct points can be selected and their feature values used as the split values. An exhaustive search would therefore require examining $B(I, J-1)$ (binomial coefficient, I objects taken $J-1$ at a time) choices of points for split values. Currently, our implementation avoids this search by using another hill-climbing search. The data is sorted, in each feature, before the main loop of the procedure begins, so that initial split values can be selected at the quantiles of the data (medians for two clusters, quartiles for four clusters, etc). At each iteration, these values are adjusted up or down by one data point (in sorted order) and the NMF recalculated. If improvements are found, the new split values are retained, otherwise not.

The time complexity of our implementation, for a fixed number of clusters, can be computed as follows. Let

I = Number of points.
J = Number of clusters.
K = Number of features.
K' = Interaction level.
R = Number of repeated searches called for ≤50.
H = Average length of the hill-climbing search.
P = Number of PRE values to compute per NMF evaluation = $B(K, K')$.
Q = Time to compute each PRE value = $I \cdot J^{K'}$.
S = Time to sort feature scores = $K \log K$

Then the time complexity of our algorithm is on the order of $R \cdot H \cdot P \cdot Q + S$. For the most common case, interaction level $K' = 1$, and with $K \leq I$, this reduces to $O(H \cdot I \cdot J \cdot K)$. The size of H is difficult to predict, and in the worst case will be exponential (J^I), but in practice we have found the search to converge quickly to a local maximum. Letting the interaction level K' increase greatly increases the complexity, because of the large number of possible interactions among features, but we have found in practice that an interaction level of one works well even with dependent features (see Section IV).

The user input to the program consists of:

- The data.
- The number of features K.
- The type, continuous or discrete, of each feature.
- The minimum and maximum number of clusters to be examined.
- Optionally, the size of the optimal feature set. Default: the total number of features.
- Optionally, the interaction level. Default: one.
- Optionally, the number of times to repeat the search. Default: no repeats.

The user can request some or all of the following output, for each number of clusters between the input minimum and maximum, or for only the number of clusters with the best NMF:

- The cluster numbers for each point.
- The NMF value for the clustering.
- The features in the optimal feature set.
- The split values for each numeric feature in the optimal feature set.
- The PRE values for each feature individually with respect to the clustering.
- Means and variances for each cluster, for numeric features.

IV. Evaluation of RIFFLE's Performance

A. Monte Carlo Studies

In this section we compare the performance of RIFFLE to k-means clustering, a standard clustering procedure with good performance on Gaussian data. We compare their ability to recover clusters in data generated by Monte Carlo methods from two or more distinct distributions. We count the number of "correct" and "incorrect" classifications by the algorithms on the basis of the distributions that actually generated the points. Since the distributions have some degree of overlap, no procedure based solely on the data could correctly determine the originating distribution for every point, and so an "optimal algorithm" was used to obtain a lower bound on accuracy. Optimal clustering was done by assigning each data point to its *most likely* originating cluster, using the known distributions that generated the data points. The optimal algorithm therefore is not a clustering algorithm but serves only to obtain a lower bound on the misclassification rate. For the RIFFLE algorithm, the interaction-level was set to one, so that features were treated as independent. For the k-means algorithm, squared Euclidean distance was used.

1) Two-Dimensional Gaussian Clusters: In our first test, we generated two-dimensional Gaussian data in two subpopulations (similar to the data used in [12]), with subpopulation means along the $x = y$ diagonal, and at several different separations in means between the two subpopulations. The separation in means ranged from one to five times the standard deviation of each subpopulation about its own mean σ. (The two subpopu-

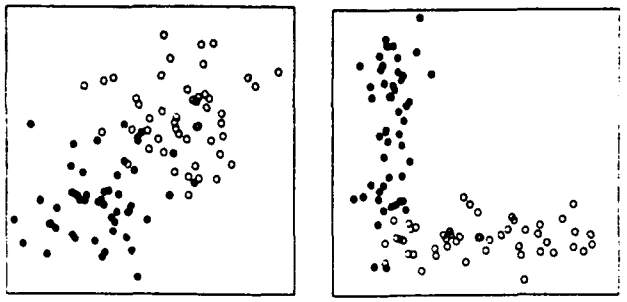


Fig. 3. Examples of synthetic two-dimensional data sets used in Monte Carlo tests. (a) Gaussian data and (b) boomerang data.

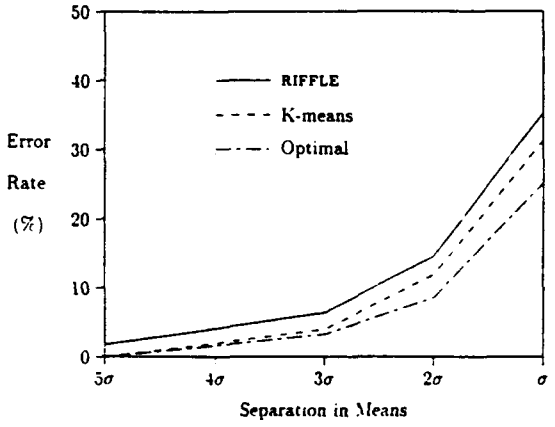


Fig. 4. Relative degradation of performance of RIFFLE, k-means, and optimal algorithms on two-dimensional Gaussian data. Errors increase for all three as the means of the two subpopulations are brought closer together. The separation in means is measured in terms of the standard deviation of each subpopulation about its mean.

lations each had the same variance.) One hundred points were generated in each experiment, with fifty in each cluster. For each parameterization the experiment was duplicated ten times, with different random number seeds, to obtain reasonable standard errors for the misclassification rates. A typical data set for this experiment is plotted in Fig. 3(a). Results for the RIFFLE and k-means algorithms, and the optimal reclassification scheme, are plotted in Fig. 4. In general, both algorithms performed well on Gaussian data.

2) Addition of Nuisance Features: A long-standing problem for many clustering algorithms [11, pp. 108–111] comes in the form of "cigar" shaped data, as illustrated in Fig. 5. Metric based clustering algorithms, which seek hyperellipsoidal clusters, typically break the cigars in half, as illustrated by the k-means clustering in Fig. 6. Clustering by RIFFLE, however, shown in Fig. 7, preserved the cigar shapes by placing more importance on the good fit of the clustering in two of the dimensions, and less importance on a poor fit in the third.

Data sets similar to the one in Fig. 5 were generated using the two-dimensional Gaussian data sets from the last section, with separation of means equal to 2σ. The cigar shape was created by introducing a third, "nuisance" feature, with values for the points randomly distributed over a range. The range of the nuisance feature varied from zero to four times the separation in means on the first two features. In Fig. 8 the performance of k-means is seen to degrade very severely as the range of nuisance noise increases. This is to be expected, since, as the nuisance feature increases in range, it dominates the other terms in the distance metric. The performance of RIFFLE, however, degrades

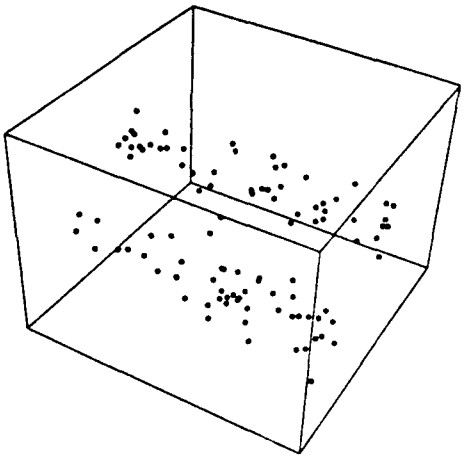


Fig. 5. Three-dimensional "cigar-shaped" data. In two dimensions the data are similar to Fig. 3(a). The points are randomly distributed in the third dimension.

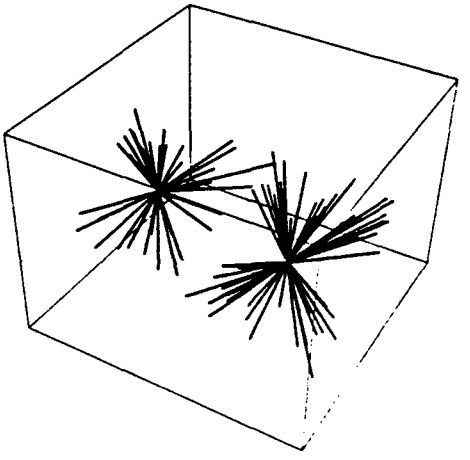


Fig. 6. Clustering of the cigar-shaped data from Fig. 5 by the k-means algorithm. Metric proximity dominates the clustering procedure, and the cigar-shaped structure is not recovered.

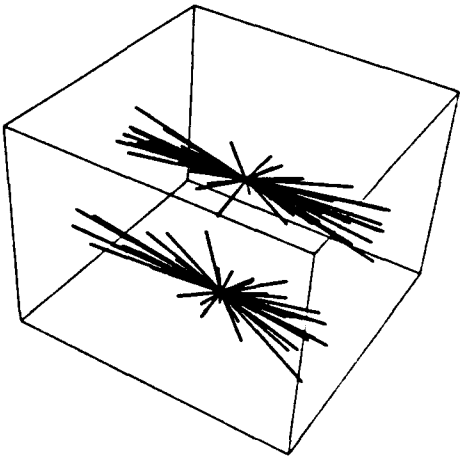


Fig. 7. Clustering of the cigar-shaped data from Fig. 5 by the RIFFLE algorithm. Because the indicated clustering fits well with two dimensions, the random third dimension is ignored.

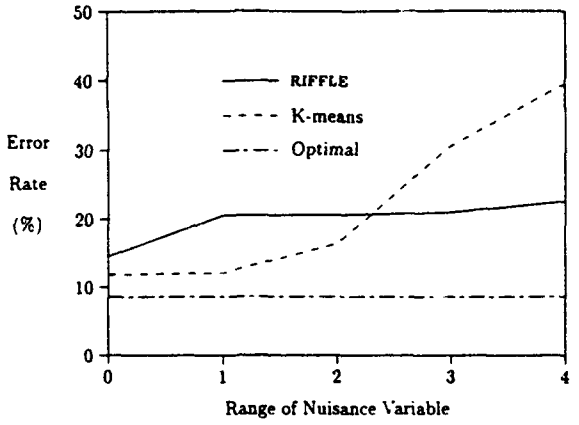


Fig. 8. Relative degradation of performance of RIFFLE, k-means, and optimal algorithms on three-dimensional, cigar-shaped data sets similar to that of Fig. 5. In these data sets the overlap in the first two dimensions was greater, and the range of the randomized third dimension was increased from zero to four times the separation in means in the first two dimensions.

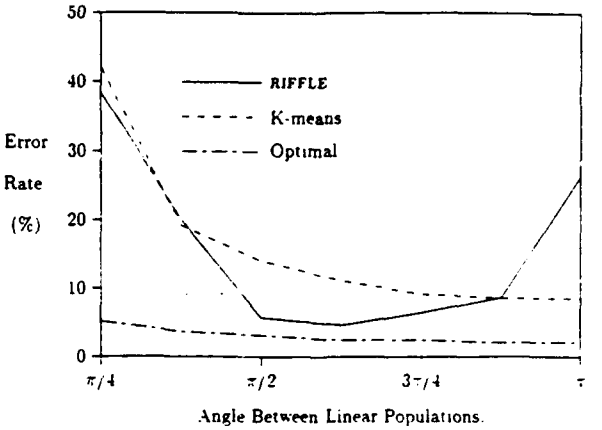


Fig. 10. Relative performance of RIFFLE, k-means, and optimal algorithms on two-dimensional boomerang data similar to that in Fig. 3(b). The angle between the two linear subpopulations varied from $\pi/4$ to π. RIFFLE's performance is equal or superior to that of k-means for most angles except the degenerate case of a straight line (π).

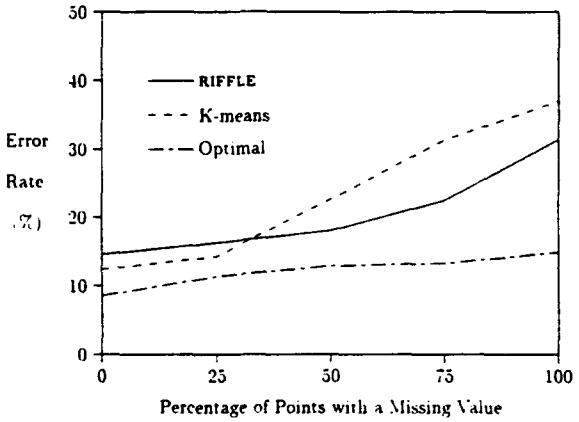


Fig. 9. Degradation of performance of RIFFLE, k-means, and optimal algorithms on two-dimensional Gaussian data similar to that in Fig. 3(a). The percentage of points which had one missing feature value was increased from zero to a hundred. The k-means algorithm required some distortion of the data in order to be usable: a substitution of the mean value for the missing values was used, and led to rapid degradation in performance. No preprocessing of the data was necessary for the RIFFLE procedure, and its degradation was less severe.

little because the fitness is nonmetric, and a good clustering in the first two features will dominate a clustering based primarily on the third feature, regardless of its range.

3) Incomplete Data: The same two-dimensional Gaussian data sets, with separation of means equal to 2σ, were also modified by taking a percentage of the points and marking one or the other of their two feature values as "missing." RIFFLE required no special treatment for missing values, since, with interaction-level one, each feature is examined independently of the others and a missing feature value is ignored in the calculation of PRE for that feature alone. However, since the standard k-means algorithm requires complete data sets, substitution of the mean value was used for missing values when running k-means. The performance of the two algorithms on this data is presented in Fig. 9. As the percentage of points with a missing value increases, the performance of k-means degrades more rapidly than RIFFLE.

4) Boomerang Data: Many data analysis situations involve clusters of points that are non-Gaussian. One common situation is when two populations represent different etiologies, but with a common origin. They tend to cluster along two different linear subspaces of the feature space, resulting in "boomerang" shaped data, such as seen in Fig. 3(b). To simulate such data, two line segments were used. The "reference" line segment was selected parallel to the x-axis. Points from one cluster were scattered uniformly along this line, with added Gaussian noise in both the x and y dimensions. The second line segment was placed at several different angles to the first, from $\pi/4$ to π, and points from the second cluster were scattered uniformly along its length, with identical Gaussian noise in x and y. A typical data set for $\pi/2$ is shown in Fig. 3(b).

Error rates for k-means and RIFFLE on these data sets are plotted in Fig. 10. For angles close to $\pi/2$ RIFFLE outperformed k-means. The reason for this is that a distance metric clustering, forced to cluster into two groups, will usually lump most of the points at the "bend" of the boomerang into the same cluster. However, in a clustering by RIFFLE, split-values close to the bend are preferred because that gives each cluster a high PRE value on at least one feature, resulting in one "horizontal" cluster and one "vertical" cluster. In Fig. 11, clusterings by k-means (a) and RIFFLE (b) for a typical boomerang data set are shown. This figure may be compared to Fig. 3(b), where the "true" subpopulations for the points are given. If there is no marked difference between the linear trends of the clusters, however, as when the angle approaches zero or π, the performance of RIFFLE breaks down.

5) Categorical Data: Categorical data was simulated with various numbers of binary features. Two subpopulations were defined by randomly choosing a single, discrete probability value prob_k for each feature, giving the probability that a sample from subpopulation one would have a "0" value on that feature. The probability that a sample from subpopulation two would have a "0" was then set at $1 - \text{prob}_k$. The experiment was repeated for a number of features varying from 3 to 8. Results for both algorithms are in Fig. 12, where it can be seen that their performances are similar.

6) Recovering the Number of Clusters: While the "true" number of clusters in a data set is an ambiguous notion, we nevertheless attempted to assess RIFFLE's performance in this area with synthetic data sets similar to those in [11]. Three data sets were generated, one with strongly clustered points, one with weakly clustered points, and one with unclustered (randomly distributed) points. Points were scattered over the unit hypercube

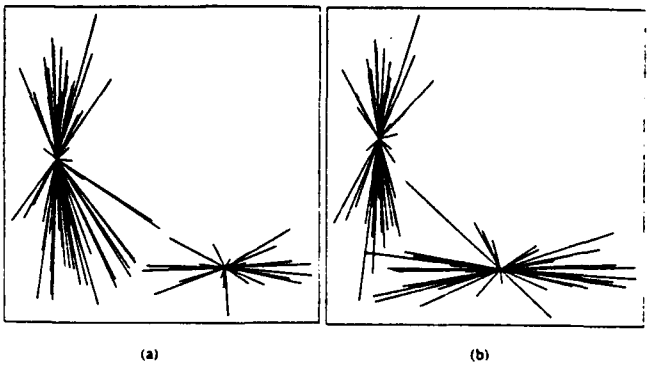


Fig. 11. Clusters generated by k-means (a) and RIFFLE (b) for boomerang data similar to that in Fig. 3(b). K-means clustering puts all the points at the "bend" in a single cluster because they are near each other in the metric. The RIFFLE clustering, however, separates the data into two subpopulations, each of which fits well with a particular dimension: one horizontal population and one vertical population.

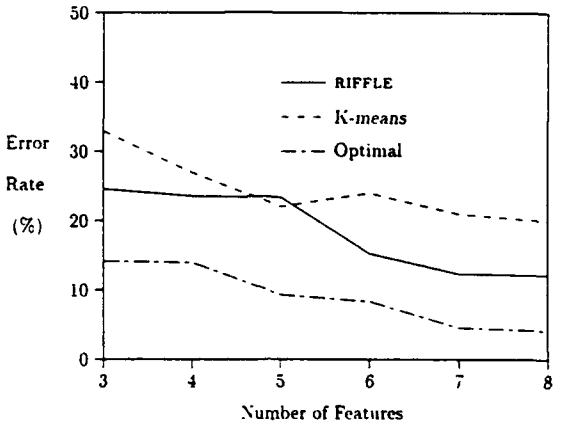


Fig. 12. Relative degradation of performance of RIFFLE, k-means, and optimal algorithms on binary categorical data. The number of binary features was varied from three to eight.

in five dimensions. For the clustered points, four subpopulation-centers were randomly selected. The strongly clustered points were normally scattered about these centers with a standard deviation of 0.01 in each dimension (no covariance); the weakly clustered points were scattered about the same centers with a standard deviation of 0.1 in each dimension. These three data sets were each clustered into two to twelve clusters by RIFFLE, and the resulting fitness values are plotted in Fig. 13. For the strongly clustered data, a clear peak is seen at the correct number, four, while for the weakly clustered data, a slight peak is still seen at the correct number. This compares well to the Davies–Bouldin index and the modified Hubert Γ index, which are plotted for similar data sets in [11, pp. 186–188]; both of these indices indicated four clusters in the strongly clustered data, but showed a slight preference for three clusters in the weakly clustered data.

In the plot for random data in Fig. 13, a tendency toward better fitness values for larger numbers of clusters can be observed. This is the well known problem of "over-fitting" a model, which plagues all data analysis situations. If necessary, a penalty for larger numbers of clusters could be introduced, perhaps along lines suggested in [13], but we have not found this necessary in practice.

B. Real World Data

1) Known Clusters: We presented two real world data sets with known properties to RIFFLE and to the k-means algorithm, to see if they could recognize the originating subpopulations. The first was Fishers's "iris" data [14], consisting of two sepal and two petal measurements from 150 irises, 50 from each of three species. We first attempted to recover the "true" number of clusters from the data. The NMF fitness values for each number of clusters from two to twelve are plotted in Fig. 14, which shows a clear peak at three. This compares favorably to the modified Hubert Γ index [15] and the fuzzy hypervolume and density indexes [16], which have been tested on the iris data, and which indicate three clusters. The Davies–Bouldin index, however, does not seem to indicate a preference for any number [15]. Using three as the correct number of clusters, we compared RIFFLE and k-means clustering. Each correctly reclassified 134 out of 150, or 88% of the irises.

The second real world data set was the "8OX" data set from [11], consisting of eight features extracted from 45 handwritten characters, 15 each of "8", "O", and "X". Fitness values for this data are also plotted in Fig. 14, but they do not reveal a clear peak. We believe this is due to the small number of data points in the 8OX data, and the fact that the clusters in the 8OX data are not well separated. Assuming, however, that each character represents a "true" cluster, we compared RIFFLE and k-means clustering on this data set. RIFFLE correctly reclassified 37 out of 45, or 82%, of the characters while k-means correctly reclassified only 30 out of 45, or 67%.

2) Unknown Clusters: In collaboration with colleagues, we are also applying RIFFLE to many ongoing problems in the analysis of real world data sets. In each case the program has

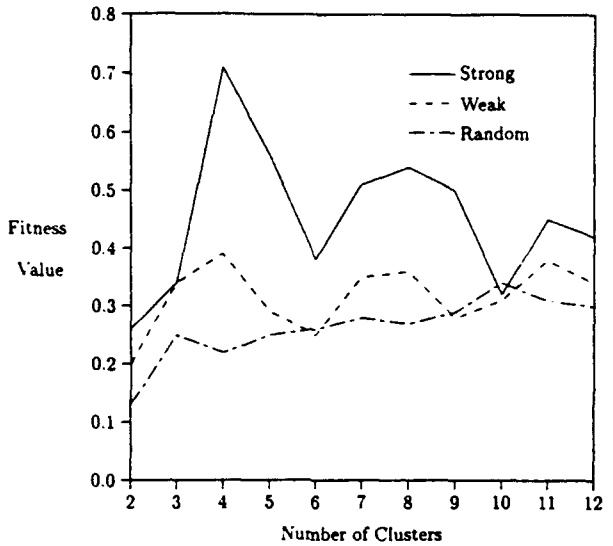


Fig. 13. Fitness values generated by RIFFLE for synthetic data with five features. Data which was strongly or weakly grouped into four subpopulations show a peak in fitness at four. Random data do not result in such a peak.

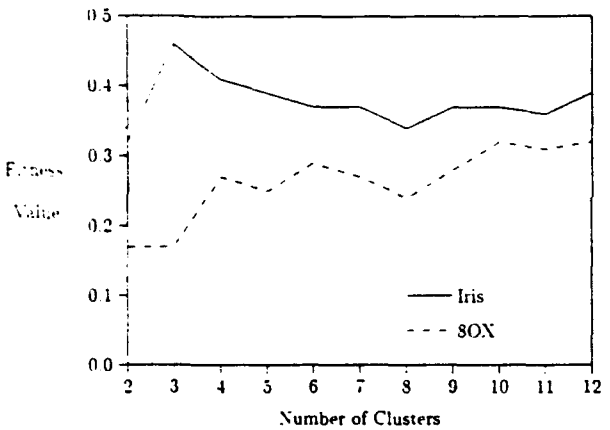


Fig. 14. Fitness values for the iris data set and the 8OX data set, for varying number of clusters, as determined by RIFFLE. For the iris data, a peak is seen at three. For the 8OX data set, no clear peak was identified, indicating that the data are not as well clustered.

created "meaningful" clusters, in some cases revealing previously unsuspected facets of the data to experts. In a year-round ecological study of a northwestern monomictic lake [17], [18], RIFFLE meaningfully clustered both the physical–chemical features and the phytoplankton species data. The physical–chemical data were separated into epilimnion, hypolimnion, and thermocline samples, even though data points were collected from three basins of the lake with quite dissimilar physical characteristics, and throughout the year. The phytoplankton samples were separated into summer versus winter samples, as these were the most dissimilar populations, with a clear break at fall turnover. Further, rare species, with low variance relative to the rest of the data set but with a high degree of association to the common algal blooms, were identified as optimal features. All other analysis tools used on the data failed to accomplish this. In another data set, gathered as part of the national acid rain survey [19] and involving hundreds of lakes, RIFFLE successfully partitioned lake samples into "impacted" and "not impacted" clusters. In a third data set, dealing with nonpoint-source pollution of an urban stream [20], RIFFLE was able to partition the samples into "polluted" and "unpolluted" clusters based solely on data involving counts of macroinvertebrates found at the sites, regardless of season. Again, other analysis tools failed to do this.

V. Conclusion

We have proposed an approach to clustering based on the principle that clusters should be selected to maximize their actual utility in predicting feature values, not ad hoc measures of similarity in feature space. We have defined a quantitative measure of this utility, called *nonmetric fitness*, which 1) is applicable to both discrete and continuous features, 2) can automatically ignore some or all noisy but irrelevant features, 3) can cluster incomplete data without assumptions about the missing values, and 4) provides some guidance in regard to the correct number of clusters. We have also implemented a clustering procedure, RIFFLE, using nonmetric fitness, and tested it on synthetic and real world data. We compared the performance of RIFFLE to *k*-means clustering and illustrated several cases where RIFFLE was superior. We are currently using RIFFLE, in collaboration with domain experts, in exploratory data analysis on real world problems, where it has proven a valuable adjunct to traditional statistical tools. We hope that our work will stimulate the creation of other clustering algorithms based on such fitness measures, as well as the use of these measures in other disciplines and data analysis tools.

References

[1] R. S. Michalski and R. E. Stepp, "Learning from observation: Conceptual clustering," in *Machine Learning, An Artificial Intelligence Approach*. Los Altos, CA: Morgan Kaufman, 1983, pp. 331–363.

[2] P. Cheeseman, J. Kelly, M. Self, J. Stutz, W. Taylor, and D. Freeman, "Autoclass: A Bayesian classification system," in *Proc. Fifth Int. Conf. Machine Learning*. Los Altos, CA: Morgan Kaufmann, 1988.

[3] L. Breiman, J. H. Freidman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.

[4] E. G. Henrichon, Jr. and K. S. Fu, "A nonparametric partitioning procedure for pattern classification," *IEEE Trans. Comput.*, vol. C-18, pp. 614–624, 1969.

[5] J. A. Sonquist, E. L. Baker, and J. N. Morgan, *Searching for Structure*, Review ed. Ann Arbor, MI: Inst. Social Res., Univ. Michigan, 1973.

[6] L. A. Goodman and W. H. Kruskal, "Measures of association for cross classifications," *J. Amer. Statist. Assoc.*, vol. 49, pp. 732–764, 1954.
[7] ——, "Measures of association for cross classifications ii: Further discussion and references," *J. Amer. Statist. Assoc.*, vol. 54, pp. 123–163, 1959.
[8] ——, "Measures of association for cross classifications iii: Approximate sampling theory," *J. Amer. Statist. Assoc.*, vol. 58, pp. 310–364, 1963.
[9] ——, "Measures of association for cross classifications iv: Simplification of asymptotic variances," *J. Amer. Statist. Assoc.*, vol. 67, pp. 415–421, 1972.
[10] L. Guttman, *An Outline of the Statistical Theory of Prediction.* New York: S.S.R.C, 1941.
[11] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data.* Englewood Cliffs, NJ: Prentice-Hall, 1988.
[12] C. K. Bayne, J. J. Beauchamp, C. L. Begovich, and V. E. Kane, "Monte Carlo comparisons of selected clustering procedures," *Pattern Recog.*, vol. 12, pp. 51–62, 1980.
[13] J. Rissanen, "Stochastic complexity and modeling," *Ann. Statist.*, vol. 14, no. 3, pp. 1080–1100, 1986.
[14] R. A. Fisher, "Multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, pp. 179–188, 1936.
[15] R. C. Dubes, "How many clusters are best?—An experiment," *Pattern Recog.*, vol. 20, no. 6, pp. 645–663, 1987.
[16] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 773–781, 1989.
[17] W. J. Ehinger, "Phytoplankton composition and temporal variation among the three basins of Lake Whatcom, Washington," Master's thesis, Western Washington Univ., Bellingham, WA, 1988.
[18] G. B. Matthews, "Clustering heterogeneous ecological data," in *Annu. Conf. Int. Soc. Ecological Modeling*, Davis, CA, 1988.
[19] R. A. Linthurst, D. H. Landers, J. M. Eilers, D. F. Brakke, W. S. Overton, E. P. Meier, and R. E. Crowe, *Characteristics of Lakes in the Eastern United State, Volume I. Population Descriptions and PhysicoChemical Relationships.* Washington, DC: U.S. Environmental Protection Agency, No. EPA/600/4-86/007a, 1986.
[20] R. A. Matthews, G. B. Matthews, and B. Hachmöller, "Ordination of benthic macroinvertebrates along a longitudinal stream gradient," in *Annu. Conf. North American Benthological Society*, Blacksburg, VA, 1990.

Geoffrey Matthews received the B.A. degree, cum laude, in philosophy from the University of California, Riverside, in 1973, and the Ph.D. degree, cum laude, in the history and philosophy of science from Indiana University in 1979.

His graduate and postdoctoral work concerned philosophical problems in the foundations of statistical physics, particularly the relations between entropy, information, causality, and the arrow of time. From 1984 to 1985 he worked at Hughes Aircraft Corporation, developing artificial intelligence applications in support of software engineering. Since 1985 he has been a Professor of Computer Science at Western Washington University, Bellingham. His research interests include pattern recognition, robust data analysis, natural language processing, and functional programming languages.

Dr. Matthews is a member of the Association for Computing Machinery and the IEEE Computer Society.

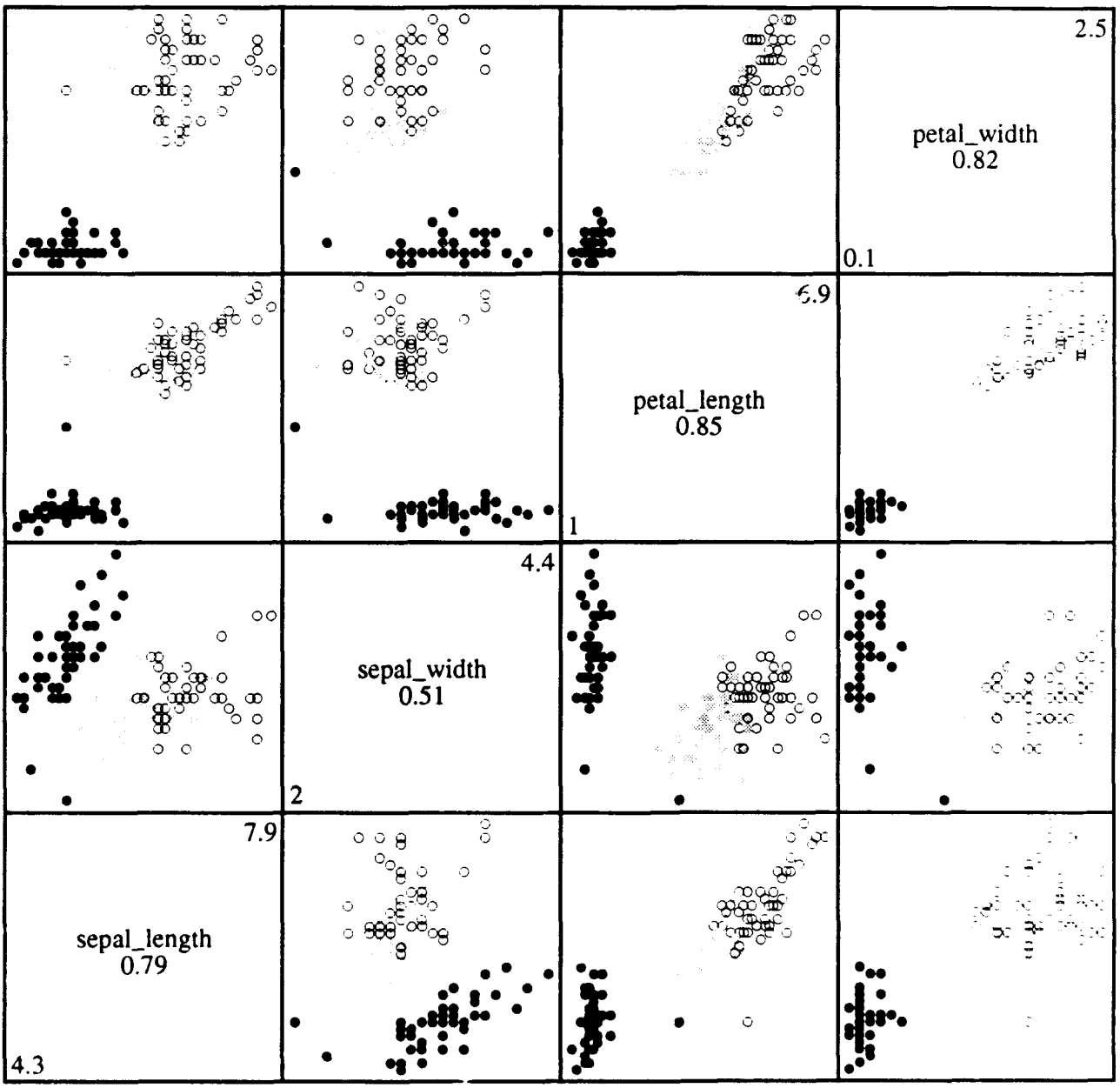
James Hearne received the B.A., M.A., and Ph.D. degrees in philosophy from the University of California, Riverside.

He has taught both Philosophy and Computer Science at several universities, including California State University at San Bernardino and Fullerton and the University of California at Riverside. He was also employed at Hughes Aircraft Company and Rockwell International where he worked on applications of artificial intelligence technology to software development environments and sonar signal processing. Since 1986 he has been an Assistant Professor of Computer Science at Western Washington University, Bellingham, where he has pursued research in exploratory data analysis and the possible worlds paradigm for distributed computing.

Statistical Ecology Minicourse

March 17 - 18, 1993
Institute of Environmental Toxicology and Chemistry

Dr. Robin A. Matthews **Dr. Geoffrey B. Matthews**



Notes

Statistical Ecology Minicourse

Dr. Robin A. Matthews
Dr. Geoffrey B. Matthews

March 17–18, 1994

Contents

Notes

Notes

1 Statistical Programming Primer

For this minicourse we will be using SPSS on a mainframe computer (called NESSIE), a few specialized statistical programs written in GWBASIC, and an ASCII text editor called EMACS. Becoming proficient in all these programs is beyond the scope of this course. Therefore, I have listed the programming commands that you will need for each exercise.

Commands that you type are written using this font: **type**.

Variables that you name are written using this font: *name*.

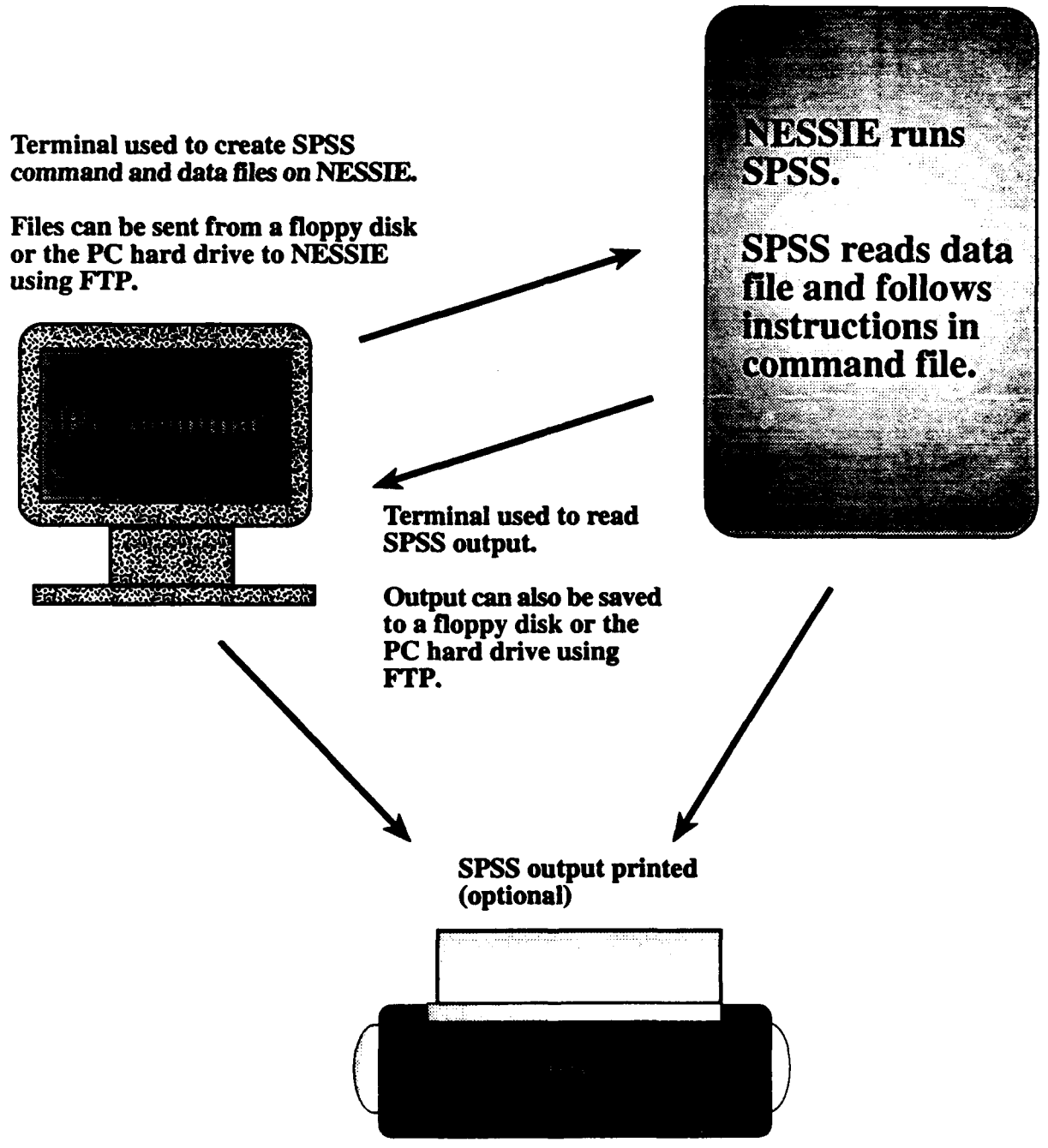
The workbook also contains special EMACS and NESSIE command subsections that list the most important commands you will be using. If you get lost, refer to the appropriate subsection, or ask your instructor.

1.1 SPSS

We will be using SPSS in "batch" mode. In batch mode, SPSS **instructions** are contained in command files (called "*filename*.com"); **data** are contained in a separate data file ("*filename*.dat"); and **output** is produced in list files ("*filename*.lis").

Here is how SPSS works: you write command and data files just like any ASCII text document. No computations are started until you submit the command file to NESSIE. When you submit the command file, NESSIE turns on SPSS, SPSS executes the command file instructions (including accessing the data file) and sends the results back to your directory on NESSIE as a .lis file. This process is illustrated in Figure 1.

Figure 1. Summary of SPSS file creation, execution, and output.



1.1.1 SPSS sample data file (iris.dat)

For most of our practice problems we will be using a simple multivariate data file containing 50 sepal and petal width and length measurements from each of three species of iris (R.A. Fisher, "The use of multiple measurements in taxonomic problems", *Annals of Eugenics* 7:179–188, 1936). Your NESSIE account contains a copy of iris.dat (listed below). Appendix A includes copies of all the output files, and Appendix B includes a summary of the patterns in the iris data.

iris.dat:

1 5.1 3.5 1.4 0.2 1 4.9 3.0 1.4 0.2 1 4.7 3.2 1.3 0.2 1 4.6 3.1 1.5 0.2 1 5.0
3.6 1.4 0.2 1 5.4 3.9 1.7 0.4 1 4.6 3.4 1.4 0.3 1 5.0 3.4 1.5 0.2 1 4.4 2.9
1.4 0.2 1 4.9 3.1 1.5 0.1 1 5.4 3.7 1.5 0.2 1 4.8 3.4 1.6 0.2 1 4.8 3.0 1.4
0.1 1 4.3 3.0 1.1 0.1 1 5.8 4.0 1.2 0.2 1 5.7 4.4 1.5 0.4 1 5.4 3.9 1.3 0.4 1
5.1 3.5 1.4 0.3 1 5.7 3.8 1.7 0.3 1 5.1 3.8 1.5 0.3 1 5.4 3.4 1.7 0.2 1 5.1
3.7 1.5 0.4 1 4.6 3.6 1.0 0.2 1 5.1 3.3 1.7 0.5 1 4.8 3.4 1.9 0.2 1 5.0 3.0
1.6 0.2 1 5.0 3.4 1.6 0.4 1 5.2 3.5 1.5 0.2 1 5.2 3.4 1.4 0.2 1 4.7 3.2 1.6
0.2 1 4.8 3.1 1.6 0.2 1 5.4 3.4 1.5 0.4 1 5.2 4.1 1.5 0.1 1 5.5 4.2 1.4 0.2 1
4.9 3.1 1.5 0.2 1 5.0 3.2 1.2 0.2 1 5.5 3.5 1.3 0.2 1 4.9 3.6 1.4 0.1 1 4.4
3.0 1.3 0.2 1 5.1 3.4 1.5 0.2 1 5.0 3.5 1.3 0.3 1 4.5 2.3 1.3 0.3 1 4.4 3.2
1.3 0.2 1 5.0 3.5 1.6 0.6 1 5.1 3.8 1.9 0.4 1 4.8 3.0 1.4 0.3 1 5.1 3.8 1.6
0.2 1 4.6 3.2 1.4 0.2 1 5.3 3.7 1.5 0.2 1 5.0 3.3 1.4 0.2 2 7.0 3.2 4.7 1.4 2
6.4 3.2 4.5 1.5 2 6.9 3.1 4.9 1.5 2 5.5 2.3 4.0 1.3 2 6.5 2.8 4.6 1.5 2 5.7
2.8 4.5 1.3 2 6.3 3.3 4.7 1.6 2 4.9 2.4 3.3 1.0 2 6.6 2.9 4.6 1.3 2 5.2 2.7
3.9 1.4 2 5.0 2.0 3.5 1.0 2 5.9 3.0 4.2 1.5 2 6.0 2.2 4.0 1.0 2 6.1 2.9 4.7
1.4 2 5.6 2.9 3.6 1.3 2 6.7 3.1 4.5 1.4 2 4.6 3.0 4.5 1.5 2 5.8 2.7 4.1 1.0 2
6.2 2.2 4.5 1.5 2 5.6 2.5 3.9 1.1 2 5.9 3.2 4.8 1.8 2 6.1 2.8 4.0 1.3 2 6.3
2.5 4.9 1.5 2 6.1 2.8 4.7 1.2 2 6.4 2.9 4.3 1.3 2 6.6 3.0 4.9 1.4 2 6.8 2.8
4.8 1.4 2 6.7 3.0 5.0 1.7 2 6.0 2.9 4.5 1.5 2 5.7 2.6 3.5 1.0 2 5.5 2.4 3.8
1.1 2 5.5 2.4 3.7 1.0 2 5.8 2.7 3.9 1.2 2 6.0 2.7 5.1 1.6 2 5.4 3.0 4.5 1.5 2
6.0 3.4 4.5 1.6 2 6.7 3.1 4.7 1.5 2 6.3 2.3 4.4 1.3 2 5.6 3.0 4.1 1.3 2 5.5
2.5 4.0 1.3 2 5.5 2.6 4.4 1.2 2 6.1 3.0 4.6 1.4 2 5.8 2.6 4.0 1.2 2 5.0 2.3
3.3 1.0 2 5.6 2.7 4.2 1.3 2 5.7 3.0 4.2 1.2 2 5.7 2.9 4.2 1.3 2 6.2 2.9 4.3
1.3 2 5.1 2.5 3.0 1.1 2 5.7 2.8 4.1 1.3 3 6.3 3.3 6.0 2.5 3 5.8 2.7 5.1 1.9 3

Notes

7.1 3.0 5.9 2.1 3 6.3 2.9 5.6 1.8 3 6.5 3.0 5.8 2.2 3 7.6 3.0 6.6 2.1 3 4.9
2.5 4.5 1.7 3 7.3 2.9 6.3 1.8 3 6.7 2.5 5.8 1.8 3 7.2 3.6 6.1 2.5 3 6.5 3.2
5.1 2.0 3 6.4 2.7 5.3 1.9 3 6.8 3.0 5.5 2.1 3 5.7 2.5 5.0 2.0 3 5.8 2.8 5.1
2.4 3 6.4 3.2 5.3 2.3 3 6.5 3.0 5.5 1.8 3 7.7 3.8 6.7 2.2 3 7.7 2.6 6.9 2.3 3
6.0 2.2 5.0 1.5 3 6.9 3.2 5.7 2.3 3 5.6 2.8 4.9 2.0 3 7.7 2.8 6.7 2.0 3 6.3
2.7 4.9 1.8 3 6.7 3.3 5.7 2.1 3 7.2 3.2 6.0 1.8 3 6.2 2.8 4.8 1.8 3 6.1 3.0
4.9 1.8 3 6.4 2.8 5.6 2.1 3 7.2 3.0 5.8 1.6 3 7.4 2.8 6.1 1.9 3 7.9 3.8 6.4
2.0 3 6.4 2.8 5.6 2.2 3 6.3 2.8 5.1 1.5 3 6.1 2.6 5.6 1.4 3 7.7 3.0 6.1 2.3 3
6.3 3.4 5.6 2.4 3 6.4 3.1 5.5 1.8 3 6.0 3.0 4.8 1.8 3 6.9 3.1 5.4 2.1 3 6.7
3.1 5.6 2.4 3 6.9 3.1 5.1 2.3 3 5.8 2.7 5.1 1.9 3 6.8 3.2 5.9 2.3 3 6.7 3.3
5.7 2.5 3 6.7 3.0 5.2 2.3 3 6.3 2.5 5.0 1.9 3 6.5 3.0 5.2 2.0 3 6.2 3.4 5.4
2.3 3 5.0 3.0 5.1 1.8

1.1.2 SPSS sample command file (iris.com)

```
$ set verify=noimage
$ spss/nobanner/out=iris.lis
file handle iris/name='iris.dat'
data list file iris free/ species sl sw pl pw
variable labels
     sl 'sepal length' /
     sw 'sepal width' /
     pl 'petal length' /
     pw 'petal width'/
set width=80
descriptives variables = sl sw pl pw
sort cases by species
split file by species
descriptives variables = sl sw pl pw
execute
finish
```

1.1.3 SPSS sample output file (iris.lis, edited)

```
11-Mar-94   SPSS RELEASE 4.1 FOR VAX/VMS
VAX                      WESTERN WASHINGTON UNIVERSITY
This software is functional through December 31, 1994.
```


Notes

```
   1  0  file handle iris/name='iris.dat'
   2  0  data list file iris free/ species sl sw pl pw
   3  0  variable labels
   4  0   sl 'sepal length' /
   5  0   sw 'sepal width' /
   6  0   pl 'petal length' /
   7  0   pw 'petal width'/
   8  0
   9  0  set width=80
  10  0
  11  descriptives variables = sl sw pl pw
  12
There are 1,498,144 bytes of memory available.
296 bytes of memory required for the DESCRIPTIVES procedure.
8 bytes have already been acquired.
288 bytes remain to be acquired.

Number of valid observations (listwise) =       150.00
                                                  Valid
Variable      Mean     Std Dev    Minimum    Maximum       N

SL            5.83         .84       4.30       7.90     150
SW            3.06         .44       2.00       4.40     150
PL            3.76        1.77       1.00       6.90     150
PW            1.20         .76        .10       2.50     150

Preceding task required .25 seconds CPU time;  .61 seconds
  13  sort cases by species
SIZE OF FILE TO BE SORTED:       150 CASES OF     40 BYTES
SORT COMPLETED SUCCESSFULLY.  FILE SIZE:          12 BLOCKS.
Preceding task required .08 seconds CPU time;  .14 seconds
  14  split file by species
  15
  16  descriptives variables = sl sw pl pw
  17
There are 1,498,848 bytes of memory available.
296 bytes of memory required for the DESCRIPTIVES procedure.
8 bytes have already been acquired.
288 bytes remain to be acquired.

SPECIES:      1.00
Number of valid observations (listwise) =        50.00
                                                   Valid
```


Notes

```
Variable      Mean     Std Dev    Minimum    Maximum        N
SL            5.01         .35       4.30       5.80       50
SW            3.43         .38       2.30       4.40       50
PL            1.46         .17       1.00       1.90       50
PW             .25         .11        .10        .60       50

SPECIES:      2.00
Number of valid observations (listwise) =        50.00
                                                    Valid
Variable      Mean     Std Dev    Minimum    Maximum        N

SL            5.92         .55       4.60       7.00       50
SW            2.77         .31       2.00       3.40       50
PL            4.27         .48       3.00       5.10       50
PW            1.33         .20       1.00       1.80       50

SPECIES:      3.00
Number of valid observations (listwise) =        50.00
                                                    Valid
Variable      Mean     Std Dev    Minimum    Maximum        N

SL            6.57         .67       4.90       7.90       50
SW            2.97         .32       2.20       3.80       50
PL            5.55         .55       4.50       6.90       50
PW            2.03         .27       1.40       2.50       50

Preceding task required .17 seconds CPU time;  .54 seconds
  18  execute
Preceding task required .02 seconds CPU time;  .02 seconds
  19  finish
    19 command lines read.
     0 errors detected.
     0 warnings issued.
     1 seconds CPU time.
     3 seconds elapsed time.
       End of job.
```


Notes

1.2 SPSS practice problem

Log on to NESSIE by double-clicking on the Telnet-Nessie icon and entering your account name and password.

Look at your NESSIE directory by typing:

```
dir↩          (↩ means hit return)
```

Look at the SPSS command file by typing the following EMACS commands:

```
emacs iris.com↩
```

Change the name of the SPSS output file from *iris.lis* to *myiris.lis*

Now look at the SPSS data file by typing:

```
^x ^f iris.dat↩
```
(^x means hold down the control key and type x)

Save the revised command file (which is in an EMACS buffer) and exit EMACS by typing:

```
^x ^s
^x ^c
```

Now submit the SPSS command file to NESSIE:

```
submit/noprint/notify iris.com↩
```

This will start SPSS and, hopefully, create a new output file called *myiris.lis*

Look at your new output, then quit EMACS and log out:

```
emacs myiris.lis↩
^x ^c
lo↩
```


Notes

EMACS QUICK REFERENCE SHEET

MEANING	KEYSTROKE
Enter EMACS	Start emacs editor and edit file named "file"emacs file <return>
Cursor Movement	By Charcter:
	UP ^p
	DOWN ^n
	LEFT ^b
	RIGHT ^f
	By Page:
	UP esc-v
	Down ^v
	LEFT ^a
	RIGHT ^e
	By Word:
	Backward esc-b
	Forward esc-f
Deleting	By Character: del ^d
	By Word: esc-del esc-d
	By Line: ^k esc-o ^k
	By Region: ^w
	To mark the begining of a region press ^@
	To yank back a deleted region, line, or word press ^y
Correcting Mistakes	To abort a command press ^g
	To reverse changes made in current editing session ^_
	To recover from a system crash esc-x recover-file
Search & Replace	Start search & replace esc-% enter string to search for <return> enter the replacement <return> The cursor will move to the first occurance of the string. Pressing the spacebar will replace. Pressing the delete key will skip over it. Pressing the ! will replace all remaining occurances.
Exiting & Saving	Save the buffer that is currently selected ^x^s
	Exit EMACS permanently ^x^c
	Attach the parent process ^x^z (This only works in VMS if EMACS was spawned)
Getting More Information	Online tutorial ^ht
	Help features accessed by ^h
	To bring up a menu of all available help ^h^h^h
	To return to document from any help screen ^x

1.4 NESSIE commands

When you are logged onto NESSIE, you will see a "$" prompt.

The most important commands for NESSIE are:

To look at the directory	`dir`↩
To copy a file	`copy` *filename* ↩
To rename a file	`rename` *filename* ↩
To delete a file	`del` *filename;** ↩
To purge old copies of files	`purge`↩
To view the job queue	`queue`↩
To check on your account quota	`quota`↩
To get on-line help	`help`↩
To submit an SPSS command file	`submit/noprint/` `notify filename.com`↩

Notes

Notes

2 Data Set Structure

2.1 ASCII data

Dat sets come in many forms. Formatting a data file so that it can be read by statistical software can be a real challenge.

Usually the easiest approach is to write a data file using an ASCII editor. The simplest ASCII data file contains only numbers and whitespace (no comments, tabs, *, bdl, slashes, dashes, etc.) in a regular, row by column rectangle. Special numeric codes (e.g., -99, -88) are used for missing data, below detection values, and non-numeric measurements (e.g. gender). A typical data file might look like this:

```
190  1  1.3  -99
210  1  2.1  -99
100  2  7.8  35
110  2  9.5  40
```

If the file is taken from a spreadsheet rather than created using an ASCII editor some manipulation is usually required to simplify the data set. The column headings need to be removed from the file, as do comments and extraneous numeric values (such as the date). Spreadsheets, word processors, and computer operating systems (DOS, Unix, VMS, MacOS, etc.) also place special characters such as `^M` or `^Z` in the data files even when they are saved as text or ASCII. These special characters (which are invisible in some text editors) must be removed prior to running SPSS or BASIC statistical programs.

Data files should be accompanied by description or readme files that describe the contents of each column. For SPSS, the command file can serve this purpose, using variable and value labels to keep track of information that is not in the data file (e.g., site descriptions, measurement units, etc.)

2.2 Common data set problems

Non-ASCII data:

As discussed above, ASCII data files are much easier to work with for statistical analyses. If the data were entered into a spreadsheet (or with a word processor) that has the ability to export ASCII text, it will probably be easier to clean up the file using the spreadsheet. If not, it may end up being easier to reenter the entire data file using an ASCII text editor. Remember that most spreadsheets insert hidden characters (tabs, end-of-lines, etc.) than may not show up on the viewing screen but will nevertheless crash statistical programs.

Below detection values:

For statistical purposes you need to decide whether to omit these data, enter a single value such one-half the difference between zero and the lowest measureable value, or try to estimate a reasonable distribution of values below the detection point.

Unbalanced data sets:

Many univariate and multivariate programs require balanced data sets. Sometimes the only choice is to leave out the unbalanced variables or average replicates so that the data are balanced.

Dependent variables:

Dependent variables increase the influence of one environmental factor on the statistical results. Sometimes dependent variables are obvious (e.g. alkalinity and bicarbonate) and a choice can be made to keep only one of the variables. Other times, the dependence is due to subtle ecological relationships (e.g. temperature and dissolved oxygen) and there is no clear resolution.

Notes

Unmeasured variables and random variables:

Variables are chosen based on whether the scientist thinks the measurement is appropriate, as well as on physical constraints (money, people, equipment, time, etc.). There is no reason to assume that we always measure the important parameters. There is no reason to assume that all measured parameters are important!

Measurement units and commensurability:

Many statistical tests, including the popular χ^2 test, are affected by changing units (e.g., from mg/L to μg/L). In addition, most parametric tests are affected by differences in the unit ranges for each parameter (e.g., pH ranges from 6.5–8.0, but bacteria range from 100,000–1,000,000).

Notes

Notes

3 Overview of Selected Statistical Procedures

3.1 Regression/Correlation

Linear regression measures the **linear** relationship between an independent variable (X) and one (or more) dependent variables (Y_i). The linear equation, $\overline{Y} = \alpha + \beta\overline{X}$, assumes that the slope (β) and intercept (α) are constant. The proportion of total variance in Y that is accounted for by the best-fit linear equation is called the coefficient of determination, r^2. The significance of a linear regression is calculated using the critical value of F (similar to ANOVA), with the null hypothesis that $\beta = 0$.

Correlation analysis measures the **linear** relationship between two variables that are not necessarily functionally dependent. The strength of this relationship is measured using the correlation coefficient, r, which is $\sqrt{r^2}$. The significance of the correlation is determined using the critical value of r, with the null hypothesis that ρ=0 (ρ is the population correlation coefficient; r is the sample correlation coefficient).

In both regression and correlation, r will be close to 1.0 if most of the points lie in a straight line. The more the points spread out, the lower the value of r. Critical values for testing the significance of r, however, are determined by sample size. (For a sample of 1000, the critical r is only 0.062!)

3.1.1 SPSS example (iriscorr.com)

```
$ set verify=noimage
$ spss/nobanner/out=iriscorr.lis
file handle iris/name='iris.dat'
data list file iris free/ species sl sw pl pw
```

```
variable labels
     sl 'sepal length' /
     sw 'sepal width' /
     pl 'petal length' /
     pw 'petal width'/
set width=80
select if(species eq 1)
************************************************************
* This subroutine calculates Pearson's r, Kendall's tau *
* and Spearman's rho correlations for iris species 1    *
************************************************************
correlate variables=sl sw pl pw
     / format = serial
nonpar corr variables=sl sw pl pw
     /format = serial
     /print=both
***********************************************
* The next subroutine plots a scatterplot of *
* sepal length vs sepal width for species 1 *
***********************************************
plot / format=regression
     / title="sepal length vs sepal width"
     / horizontal="sepal width"
     / vertical="sepal length"
     / plot sl with sw
************************************************************
* The next subroutine calculates regression statistics *
* on sepal length vs sepal width for species 1         *
************************************************************
regression / variables = sl sw
     / dependent = sl
     / method = enter sw
     / scatterplot (*sresid *pred)
     / residuals = histogram (sresid)
************************************************************
* The last subroutine calculates multiple regression for *
* all flower measurements for species 1 only             *
************************************************************
regression
     / variables = sl sw pl pw
     / dependent = sl
     / method = stepwise
execute
finish
```


Notes

3.2 ANOVA/MANOVA

Analysis of variance procedures (t test, ANOVA, and MANOVA) all use ratios of within-group variance to total variance to test for significant differences between groups. The null hypothesis is that the population means are **not** significantly different. The significance is determined using the F statistic, which is the ratio of the averaged group variance to the non-group variance. The F statistic is compared to a table of critical F values; F statistics greater than the critical value result in rejection of the null hypothesis. The decision to accept or reject the null hypothesis carries a probability (p) of committing a Type I error, which is rejection of the null hypothesis when it is actually true.

ANOVA is used to test whether **any** of the groups are significantly different. It is often used in conjunction with a multiple range test to determine which groups are different from the others. The multiple range test should not be used if the ANOVA results are not significant.

The three most important assumptions for using ANOVA are that the samples were collected randomly, the results are distributed normally, and the variances are homogeneous. These three assumptions are rarely met in ecological data, and most uses of ANOVA rely heavily its ability to perform despite departures from normality and homogeneity.

Notes

3.2.1 SPSS example (irisanov.com)

```
$ set verify=noimage
$ spss/nobanner/out=irisanova.lis
file handle iris/name='iris.dat'
data list file iris free/ species sl sw pl pw
variable labels
     sl 'sepal length' /
     sw 'sepal width' /
     pl 'petal length' /
     pw 'petal width'/
set width=80
*****************************************************************
* This calculates ANOVA and multiple ranges for sepal length,*
* lists descriptive statistics, and test homogeneity of var. *
*****************************************************************
oneway sl by species(1,3)
     / ranges = duncan
     / ranges = snk
     / statistics = all
*******************************************************
* This calculates a nonparametric version of ANOVA *
*******************************************************
npar tests k-w = sl by species(1,3)
*************************************************************
* This calculates multiple analysis of variance for    *
* all four flower measurements by species              *
*************************************************************
manova sl sw pl pw
     by species(1,3)
     / print=homogeneity(all)
     / power
     / print=signif(efsize)
     / cinterval=multivariate(wilks)
execute
finish
```


Notes

3.3 Species Diversity

Species diversity is composed of two factors: the number of species in the population (species richness) and the distribution of individuals among the species in a population (species equitability). Changes in species diversity are often used to look for the effects of pollutants or disturbances at a site, using the assumption that "clean" sites will have many different species (high species richness), and a large number of moderately abundant species (high equitability). Polluted sites, by contrast, would have low species richness and a few very abundant species (low equitability).

All of the species richness, diversity, and equitability indices are influenced by the level of effort that goes into sampling, counting, and identifying the species. Many of these indices are particularly affected by the lowest taxonomic level used for separating taxa (e.g., species, genera, family, etc.)

Some commonly used species richness, species diversity, and equitability indices are listed below. The numbers correspond to the sample problem results in the next section.

Shannon's index (entropy) = $H' = -\sum_{i=1}^{S}(\frac{n_i}{n})\ln(\frac{n_i}{n})$

Simpson's index = $\lambda = -\sum_{i=1}^{S}\frac{n_i(n_i-1)}{n(n-1)}$

N0 = S (total # of species in sample)

N1 = $e^{H'}$ (# of abundant species in sample)

N2 = $1/\lambda$ (# of very abundant species in sample)

E1 (Pielou's J') = $\frac{H'}{\ln(S)}$ (ratio of H' to maximum H')

E2 (Sheldon's index) = $\frac{e^{H'}}{S}$

Notes

E3 (Heip's index) = $\frac{e^{H'}-1}{S-1}$

E4 (Hill's index) = $\frac{1/\lambda}{e^{H'}}$

E5 (modified Hill's index) = $\frac{(1/\lambda)-1}{e^{H'}-1}$

3.3.1 GWBASIC example (divers.lis)

This example is from the program SPDIVERS.BAS (GWBASIC). This program, and many more, are supplied with the textbook *Statistical Ecology* by John A. Ludwig and James F. Reynolds (John Wiley & Sons, New York, 1988). The program calculates a number of different species richness, diversity, and equitability indices. It is being run on the following two data sets:

Data Set 1

Species	Abundance
A	500
B	300
C	200

Data Set 2

Species	Abundance
A	500
B	299
C	200
D	1

Notes

SPDIVERS.BAS edited output file:

```
          THIS PROGRAM COMPUTES:
             1. RICHNESS INDICES
                   Margalef  (R1, Eq. 8.1)
                   Menhinick (R2, Eq. 8.2)
             2. DIVERSITY INDICES
                   Hill's Numbers (N0,N1,N2, Eq. 8.5)
                   Simpson's Index (Lambda, Eq. 8.7)
                   Shannon's Index (H', Eq. 8.9)
            3. EVENNESS INDICES
                   E1-E5, Eqs. 8.11-8.15

ENTER abundance data for each species
SPECIES__ 1
 ? 500
SPECIES__ 2
 ? 300
SPECIES__ 3
 ? 200

RICHNESS
--------
     N0 =  3
     R1 =  .2895297
     R2 =  9.486833E-02

DIVERSITY
---------
 LAMBDA =  .3793794
     H' =  1.029653
     N1 =  2.800094
     N2 =  2.635884

EVENNESS
--------
     E1 =  .9372306
     E2 =  .9333647
     E3 =  .9000471
     E4 =  .9413555
     E5 =  .9087769
```


Notes

```
ENTER abundance data for each species
SPECIES__ 1
 ? 500
SPECIES__ 2
 ? 299
SPECIES__ 3
 ? 200
SPECIES__ 4
 ? 1

RICHNESS
--------
     N0 =  4
     R1 =  .4342945
     R2 =  .1264911

DIVERSITY
---------
 LAMBDA =  .3787808
     H' =  1.036355
     N1 =  2.818924
     N2 =  2.64005

EVENNESS
--------
     E1 =  .7475723
     E2 =  .7047309
     E3 =  .6063078
     E4 =  .9365452
     E5 =  .9016594
```


Notes

3.4 Cluster Analysis

Cluster analysis is a multivariate technique used to find patterns of similarity in groups of samples. Our examples will use agglomerative clustering (the clusters start small and get bigger by adding "close" samples) with variations on how distance is measured between clusters and how the new cluster center is calculated after two samples are joined.

Most clustering programs are exploratory (they help find patterns) but not confirmatory (they don't test the significance of a pattern). Also, clustering programs must use all of the measured variables in the data set, so the inclusion of random variables or outliers can have serious consequences.

We will use two distance measures in the SPSS example:

Squared Euclidean distance $= \sum_i (x_i - y_i)^2$

Cosine distance $= \frac{\sum_i (x_i y_i)}{\sqrt{\sum_i x_i^2 \times \sum_i y_i^2}}$

Notes

3.4.1 SPSS example (irisclust.com)

```
$ set verify=noimage
$ spss/nobanner/out=iriscluster.lis
file handle iris/name='iris.dat'
data list file iris free/ species sl sw pl pw
variable labels
     sl 'sepal length' / sw 'sepal width' /
     pl 'petal length' / pw 'petal width'/
set width=80
*************************************************
* This clusters the iris data using SE distance *
* with the centroid clustering method           *
*************************************************
cluster sl sw pl pw
     / plot = dendrogram
     / method = centroid
     / measure = seuclid
*********************************************************
* This clusters the iris data using cosine distance     *
* with the average linkage clustering method            *
*********************************************************
cluster sl sw pl pw
     / plot = dendrogram
     / method = baverage
     / measure = cosine
execute
finish
```


Notes

3.5 Factor Analysis (PCA, COA, etc.)

Factor analysis is a widely used multivariate technique that attempts to identify new variables (from a composite of the old variables) that can be used to separate groups. It starts by computing a matrix of the "similarity" between groups (PCA typically uses a correlation matrix). The similarity matrix is usually standardized to center the data around variable means of 0 and variances of 1. The second step is to find the **linear** combination of weighted variable scores that contains the greatest amount of variance (factor extraction). The first principal component is the combination that accounts for the greatest amount of variance. The second principal component is the combination that accounts for the next greatest amount to variance and is uncorrelated with the first principal component. There are as many principal components as there are variables in the data, and if you use all of them, you account for all of the variance. However, the idea is to find a large amount of variance in the first few principal components, then analyse the factor scores to see which variables contributed most to each component.

Correspondence analysis is similar in concept; the major differences occur in how the data matrix is transformed.

As with all of the other multivariate, parametric statistics, factor analysis is very much influenced by random variables, outliers, and subtle changes in the grouping of the data. It appears to work best on data that have strong linear patterns in a few frequently measured variables.

Notes

3.5.1 SPSS example (irispca.com)

Some statistical programs allow you to run PCA in either sampling unit (SU) mode or species (variables) mode (see GWBASIC example). However, the only "easy" SPSS approach is to ordinate the variables.

```
$ set verify=noimage
$ spss/nobanner/out=irispca.lis
file handle iris/name='iris.dat'
data list file iris free/ species sl sw pl pw
variable labels
     sl 'sepal length' /
     sw 'sepal width' /
     pl 'petal length' /
     pw 'petal width'/
set width=80
*********************************************
* This computes PCA and plots PCI and PCII *
*********************************************
factor var=sl sw pl pw
     / criteria = factors(4)
     / plot = rotation(1,2)
execute
finish
```


Notes

3.5.2 GWBASIC example (rotate.lis, edited)

The GWBASIC program PCA.BAS (*Statistical Ecology* by Ludwig and Reynolds, 1988) allows you to select either species ordination or SU ordination, or both. However, because it runs on a PC, it can't handle the entire iris data set, so the example shown below is for a subset of 20 SUs from each iris species (named rotate.dat in Disk-1).

The choices are a bit confusing in this program. We will perform Option #1 (SU ordination), using the different petal and sepal measurements are "species". All references to species 1–4 translate to (1) sepal length; (2) sepal width; (3) petal length; and (4) petal width. The program doesn't plot the output, but this can be done using any standard plotting program.

PCA.BAS edited output file:

```
   STATISTICAL ECOLOGY:  A PRIMER ON METHODS AND COMPUTING
   I N T E R A C T I V E     B A S I C     P R O G R A M
                    P C A . B A S
   ------------------------------------------------------
   This PROGRAM COMPUTES a PRINCIPAL COMPONENTS ANALYSIS
   for THREE COMPONENTS based on SPECIES CORRELATIONS
   ------------------------------------------------------
   OPTIONS included are:
        Option # 1.  Sampling Unit (SU) Ordination
        Option # 2.  Species Ordination
        Option # 3.  BOTH SU and SPECIES Ordination
   ------------------------------------------------------
INPUT your CHOICE of Options (1-3) ? 1

- - - - - - - - - - PART I.  DATA ENTRY - - - - - - - - - - -
INPUT the NUMBER of SAMPLING UNITS (SUs) ? 60
INPUT the NUMBER of SPECIES ? 4
Specify name of DATA File (e.g., PCA.DAT) ? rotate.dat
Specify DISK DRIVE where located: A, B, C, etc. ? b

- - -  PART II.  PRINCIPAL COMPONENTS ANALYSIS - - -
```


Notes

R MATRIX (SPECIES CORRELATIONS) - UPPER TRIANGLE
Correlation between Species 1) and Species
2) = -0.047 3) = 0.825 4) = 0.763
Correlation between Species 2) and Species
3) = -0.436 4) = -0.372
Correlation between Species 3) and Species
4) = 0.967

Summary of Eigenanalysis:

EIGENVALUE		PERCENT OF TRACE	ACCUMULATED % of TRACE
1 =	2.846	71.1%	71.1%
2 =	0.963	24.1%	95.2%
3 =	0.175	4.4%	99.6%
4 =	0.016	0.4%	100.0%

EIGENVECTOR 1 = 0.506 -0.267 0.589 0.571
EIGENVECTOR 2 = 0.436 0.899 0.001 0.034
EIGENVECTOR 3 = -0.706 0.319 0.158 0.612
EIGENVECTOR 4 = -0.235 0.134 0.793 -0.546

SAMPLING UNIT Coordinates on the 1st 3 Principal Components

	COMPONENTS		
SU	I	II	III
1	-0.283	0.052	-0.015
2	-0.262	-0.079	-0.036
3	-0.296	-0.044	0.001
4	-0.288	-0.075	0.005
5	-0.297	0.070	0.004
6	-0.257	0.168	0.011
7	-0.303	-0.003	0.039
8	-0.279	0.022	-0.012
9	-0.293	-0.135	0.008
10	-0.275	-0.056	-0.037
11	-0.270	0.119	-0.028
12	-0.290	0.009	0.010
13	-0.280	-0.086	-0.036
14	-0.330	-0.118	0.013
15	-0.274	0.216	-0.048
16	-0.278	0.305	0.020
17	-0.273	0.168	0.007
18	-0.273	0.053	-0.005

Notes

19	-0.237	0.163	-0.039
20	-0.290	0.124	0.022
21	0.134	0.111	-0.078
22	0.090	0.073	-0.007
23	0.151	0.081	-0.064
24	0.046	-0.200	-0.015
25	0.130	-0.016	-0.050
26	0.047	-0.068	0.011
27	0.094	0.090	0.024
28	-0.064	-0.217	0.017
29	0.111	0.013	-0.073
30	0.001	-0.124	0.059
31	-0.020	-0.305	-0.025
32	0.054	-0.007	0.025
33	0.062	-0.193	-0.107
34	0.088	-0.018	-0.009
35	-0.005	-0.051	0.020
36	0.110	0.068	-0.057
37	-0.030	-0.091	0.164
38	0.016	-0.087	-0.043
39	0.143	-0.177	-0.071
40	0.016	-0.147	-0.031
41	0.234	0.096	0.131
42	0.144	-0.082	0.061
43	0.273	0.074	-0.020
44	0.179	-0.003	0.021
45	0.233	0.036	0.052
46	0.340	0.106	-0.064
47	0.046	-0.189	0.111
48	0.283	0.061	-0.076
49	0.245	-0.072	-0.052
50	0.285	0.225	0.063
51	0.171	0.082	0.040
52	0.197	-0.044	0.001
53	0.234	0.054	0.007
54	0.156	-0.136	0.064
55	0.185	-0.056	0.121
56	0.200	0.077	0.084
57	0.182	0.033	0.007
58	0.305	0.303	0.004
59	0.407	0.019	-0.085
60	0.152	-0.190	-0.044

Notes

3.6 Discriminant Analysis

Discriminant analysis is similar to PCA in that it looks for a linear combination of variables that can be used to separate groups. One important feature is that discriminant analysis selects the best variables for separating predetermined groups. Random variables should be less of a problem than in multivariate procedures that use all variables (MANOVA, clustering, PCA), but group definition can be tricky. The variables that you include in the analysis are used to generate discriminant functions, which are like a set of rules that help decide which group a sample comes from based on its variable measurements. You can use discriminant analysis to predict group membership for unknown samples or to check the group assignments of known samples (this shows whether the "rules" can reproduce the actual group memberships).

Discriminant analysis tends to overfit the data set, so it is usually better at confirming group memberships for the known samples than predicting group memberships of unknown samples.

3.6.1 SPSS example (irisdisc.com)

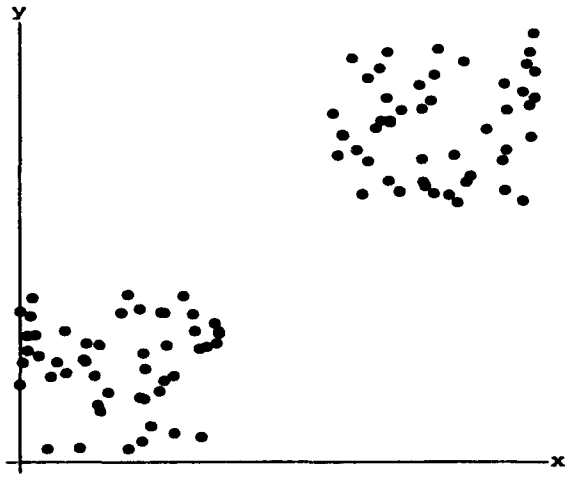
```
$ set verify=noimage
$ spss/nobanner/out=irisdisc.lis
file handle iris/name='iris.dat'
data list file iris free/ species sl sw pl pw
variable labels
     sl 'sepal length' / sw 'sepal width' /
     pl 'petal length' / pw 'petal width'/
set width=80
discriminant groups=species(1,3)
     / variables sl sw pl pw / method = wilks
     / statistics = all / plot
execute
finish
```


Notes

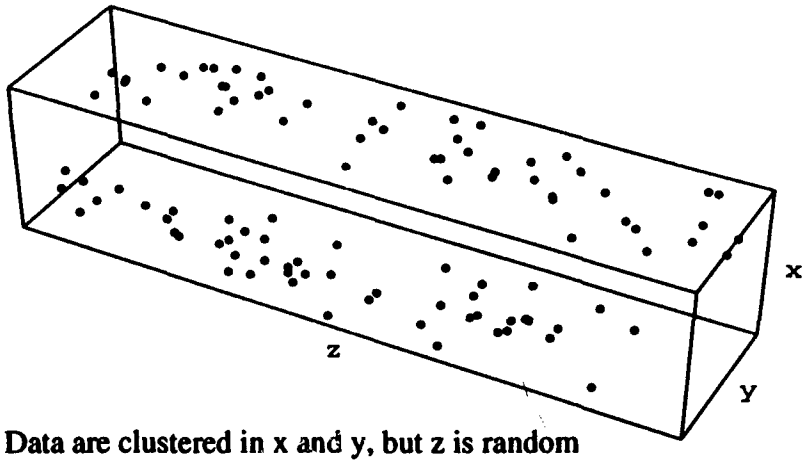
3.7 Nonmetric Clustering and Association Analysis

Nonmetric clustering and association analysis is fundamentally different than most multivariate procedures in two ways: 1) it does not use a multivariate distance measure to define similarity between groups; and 2) it ranks the variable measurements into "large" and "small" categories rather than using direct counts or measurements. NCAA follows the strategy that the best clusters are those that have the most features in common. This strategy is illustrated in Figure 2. Here, hypothetical data are shown that are strongly clustered into two groups in the x and y dimensions (Figure 2a), but the third feature, z is completely random (Figure 2b). Traditional clustering programs, which use all three features, are misled by the noise in z to the extent that they generate two completely incorrect clusters that split the data along the z axis (Figure 2c). Nonmetric clustering recognizes that z cannot improve upon the separation provided by x and y, so it clusters (correctly) on x and y, ignoring z (Figure 2d).

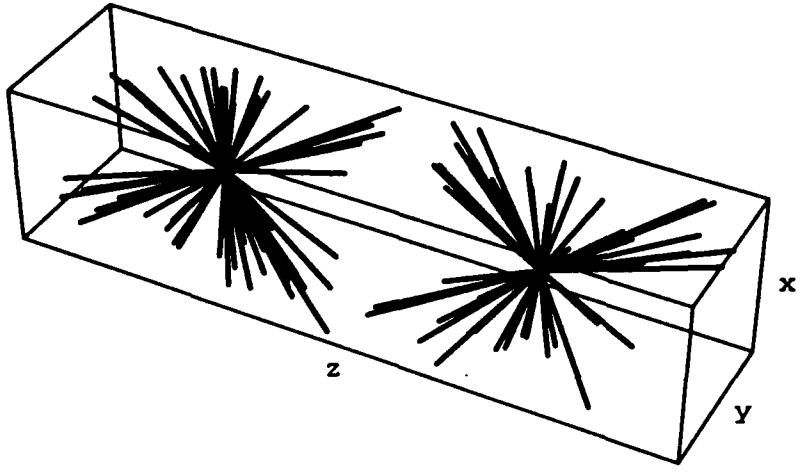
Notes



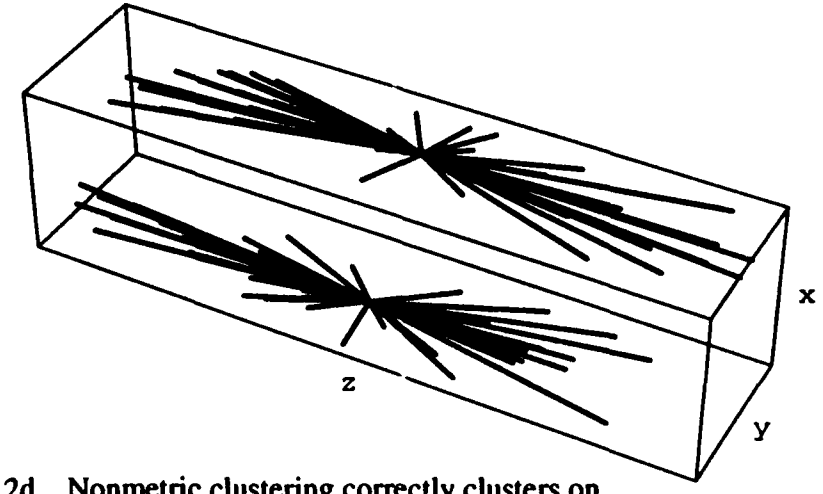
2a. Data are clustered in x and y



2b. Data are clustered in x and y, but z is random



2c. Metric clustering incorrectly clusters on z



2d. Nonmetric clustering correctly clusters on x and y, ignoring z

Figure 2. Comparison of metric and nonmetric clustering.

3.7.1 RIFFLE demonstration (iris.dat)

```
    R I F F L E
    Version 1.03  Wed Sep  8 07:36:34 PDT 1993
Data file: iris.dat
Clustering 150 points in 5 attributes into 3 clusters using
4 significant attributes and 2 retries.
Attribute      Qual Rnk1      Val1 Rnk2      Val2
  sepal_length 0.79  51       6.30  96       5.50
   sepal_width 0.51  50       3.20  99       2.90
  petal_length 0.85  53       4.80 100       3.00
   petal_width 0.82  50       1.60 100       1.00
Average Qual:  0.74

Contingency table:
              clusters
           50    0    0  |  50
  groups    1   39   10  |  50
            0    9   41  |  50
          ---------------
           51   48   51
Association analysis (chi-square significance): 1.000000
============================================================
```

Notes

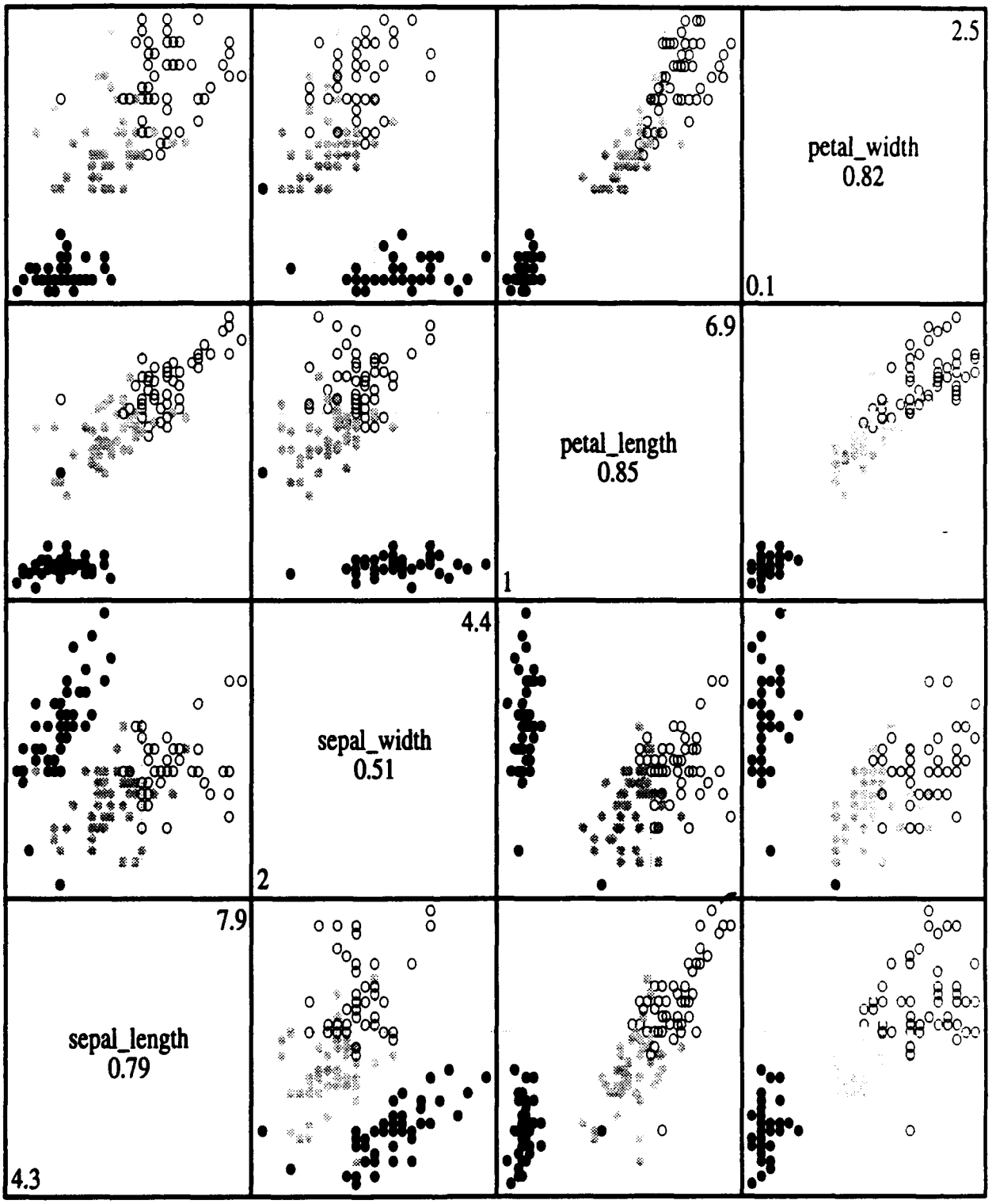


Figure 3. Scatterplot matrix showing iris data groups and Riffle clusters.

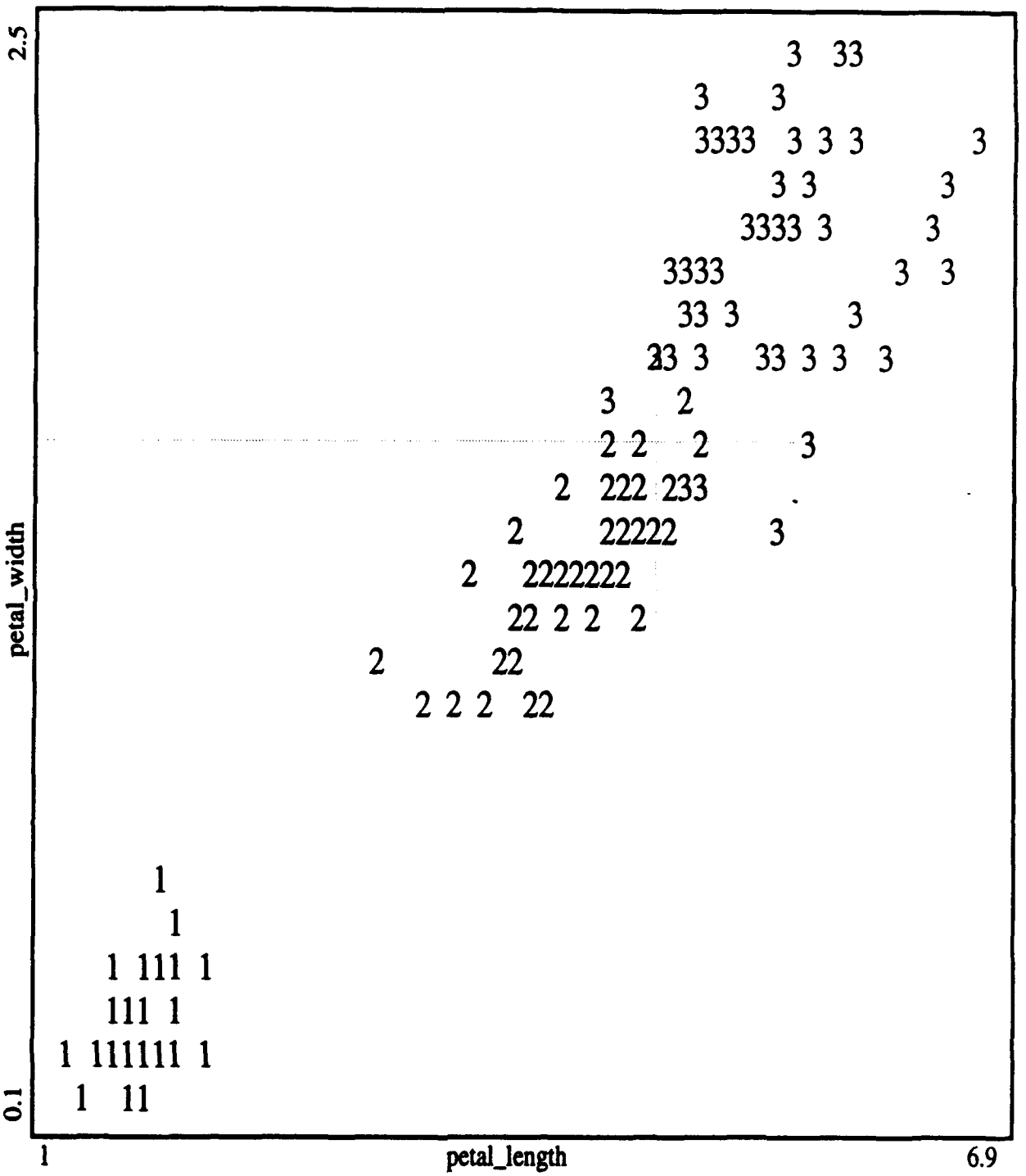


Figure 4. Scatterplot of "best view" Riffle cluster assignments.
Numbers show iris species (1–3)

Appendix A. SPSS Output Files

```
 11-Mar-94   SPSS RELEASE 4.1 FOR VAX/VMS
Page   1
15:36:11    WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

VAX                      WESTERN WASHINGTON UNIVERSITY     License Number 20077
This software is functional through December 31, 1994.

   1  0  file handle iris/name='iris.dat'
   2  0  data list file iris free/ species sl sw pl pw
   3  0  variable labels
   4  0       sl 'sepal length' /
   5  0       sw 'sepal width' /
   6  0       pl 'petal length' /
   7  0       pw 'petal width'/
   8  0
   9  0  set width=80
  10  0
  11  descriptives variables = sl sw pl pw
  12

There are 1,498,144 bytes of memory available.


296 bytes of memory required for the DESCRIPTIVES procedure.
8 bytes have already been acquired.
288 bytes remain to be acquired.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page   2
15:36:12  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


Number of valid observations (listwise) =       150.00

                                                      Valid
Variable      Mean     Std Dev   Minimum   Maximum       N  Label

SL            5.83         .84      4.30      7.90    150  sepal length
SW            3.06         .44      2.00      4.40    150  sepal width
PL            3.76        1.77      1.00      6.90    150  petal length
PW            1.20         .76       .10      2.50    150  petal width
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page   3
15:36:12  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


Preceding task required .25 seconds CPU time;  .61 seconds elapsed.

  13  sort cases by species

SIZE OF FILE TO BE SORTED:       150 CASES OF     40 BYTES EACH.
SORT COMPLETED SUCCESSFULLY.  FILE SIZE:          12 BLOCKS.

Preceding task required .08 seconds CPU time;  .14 seconds elapsed.

  14  split file by species
  15
  16  descriptives variables = sl sw pl pw
  17

There are 1,498,848 bytes of memory available.


296 bytes of memory required for the DESCRIPTIVES procedure.
8 bytes have already been acquired.
288 bytes remain to be acquired.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page   4
15:36:13  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4
```

```
SPECIES:      1.00

Number of valid observations (listwise) =       50.00

                                                    Valid
Variable      Mean     Std Dev   Minimum   Maximum      N  Label

SL            5.01         .35      4.30      5.80     50  sepal length
SW            3.43         .38      2.30      4.40     50  sepal width
PL            1.46         .17      1.00      1.90     50  petal length
PW             .25         .11       .10       .60     50  petal width
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   5
15:36:13  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

SPECIES:      2.00

Number of valid observations (listwise) =       50.00

                                                    Valid
Variable      Mean     Std Dev   Minimum   Maximum      N  Label

SL            5.92         .55      4.60      7.00     50  sepal length
SW            2.77         .31      2.00      3.40     50  sepal width
PL            4.27         .48      3.00      5.10     50  petal length
PW            1.33         .20      1.00      1.80     50  petal width
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   6
15:36:13  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

SPECIES:      3.00

Number of valid observations (listwise) =       50.00

                                                    Valid
Variable      Mean     Std Dev   Minimum   Maximum      N  Label

SL            6.57         .67      4.90      7.90     50  sepal length
SW            2.97         .32      2.20      3.80     50  sepal width
PL            5.55         .55      4.50      6.90     50  petal length
PW            2.03         .27      1.40      2.50     50  petal width
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   7
15:36:13  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


Preceding task required .17 seconds CPU time;  .54 seconds elapsed.

  18  execute

Preceding task required .02 seconds CPU time;  .02 seconds elapsed.

  19  finish

    19 command lines read.
     0 errors detected.
     0 warnings issued.
     1 seconds CPU time.
     3 seconds elapsed time.
       End of job.
```

```
 14-Mar-94   SPSS RELEASE 4.1 FOR VAX/VMS
Page   1
17:13:28    WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

VAX                      WESTERN WASHINGTON UNIVERSITY     License Number 20077
This software is functional through December 31, 1994.

   1  0  file handle iris/name='iris.dat'
   2  0  data list file iris free/ species sl sw pl pw
   3  0  variable labels
   4  0       sl 'sepal length' /
   5  0       sw 'sepal width' /
   6  0       pl 'petal length' /
   7  0       pw 'petal width'/
   8  0
   9  0  set width=80
  10  select if(species eq 1)
  11
  12  ************************************************************
  13  * This subroutine calculates Pearson's r, Kendall's tau *
  14  * and Spearman's rho correlations for iris species 1     *
  15  ************************************************************
  16  correlate variables=sl sw pl pw
  17          / format = serial
  18

PEARSON CORR problem requires 352 bytes of workspace.
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   2
17:13:29  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

                    - -  Correlation Coefficients  - -

Variable            Variable            Variable            Variable
Pair                Pair                Pair                Pair
--------            --------            --------            --------

SL          .7425   SL          .2672   SL          .2781   SW          .1777
with     N(   50)   with     N(   50)   with     N(   50)   with     N(   50)
SW       Sig .000   PL       Sig .061   PW       Sig .051   PL       Sig .217

SW          .2328   PL          .3316
with     N(   50)   with     N(   50)
PW       Sig .104   PW       Sig .019

Sig is 2-tailed, '.' is printed if a coefficient cannot be computed.
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   3
17:13:29  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


Preceding task required .23 seconds CPU time;  .32 seconds elapsed.

  19  nonpar corr variables=sl sw pl pw
  20          /format = serial
  21          /print=both
  22
  23
  24  ************************************************
  25  * The next subroutine plots a scatterplot of *
  26  * sepal length vs sepal width for species 1 *
  27  ************************************************

There are 1,498,944 bytes of memory available.


***** WORKSPACE ALLOWS FOR   40154 CASES FOR NONPARAMETRIC CORRELATION PROBLEM *****

 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   4
```

```
17:13:30  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

- - - - K E N D A L L   C O R R E L A T I O N   C O E F F I C I E N T S - - -

VARIABLE          VARIABLE          VARIABLE          VARIABLE
PAIR              PAIR              PAIR              PAIR
--------          --------          --------          --------

SL          .5973 SL          .2173 SL          .2311 SW          .1426
WITH     N(   50) WITH     N(   50) WITH     N(   50) WITH     N(   50)
SW       SIG .000 PL       SIG .022 PW       SIG .021 PL       SIG .094

SW          .2343 PL          .2217
WITH     N(   50) WITH     N(   50)
PW       SIG .020 PW       SIG .030

. . * IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                    Page   5
17:13:30  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

- - - S P E A R M A N   C O R R E L A T I O N   C O E F F I C I E N T S - - -

VARIABLE          VARIABLE          VARIABLE          VARIABLE
PAIR              PAIR              PAIR              PAIR
--------          --------          --------          --------

SL          .7553 SL          .2789 SL          .2995 SW          .1799
WITH     N(   50) WITH     N(   50) WITH     N(   50) WITH     N(   50)
SW       SIG .000 PL       SIG .025 PW       SIG .017 PL       SIG .106

SW          .2865 PL          .2711
WITH     N(   50) WITH     N(   50)
PW       SIG .022 PW       SIG .028

. . * IS PRINTED IF A COEFFICIENT CANNOT BE COMPUTED.
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                    Page   6
17:13:30  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

Preceding task required .24 seconds CPU time;  .40 seconds elapsed.

  28  plot / format=regression
  29           / title='sepal length vs sepal width'
  30           / horizontal='sepal width'
  31           / vertical='sepal length'
  32           / plot sl with sw
  33
  34
  35  ************************************************************
  36  * The next subroutine calculates regression statistics     *
  37  * on sepal length vs sepal width for species 1             *
  38  ************************************************************

There are 1,497,152 bytes of memory available.

PLOT requires 14984 bytes of workspace for execution.
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                    Page   7
17:13:30  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4
```

* * * * * * * * * * * * * * * * * * P L O T * * * * * * * * * * * * * * * * * *

Data Information

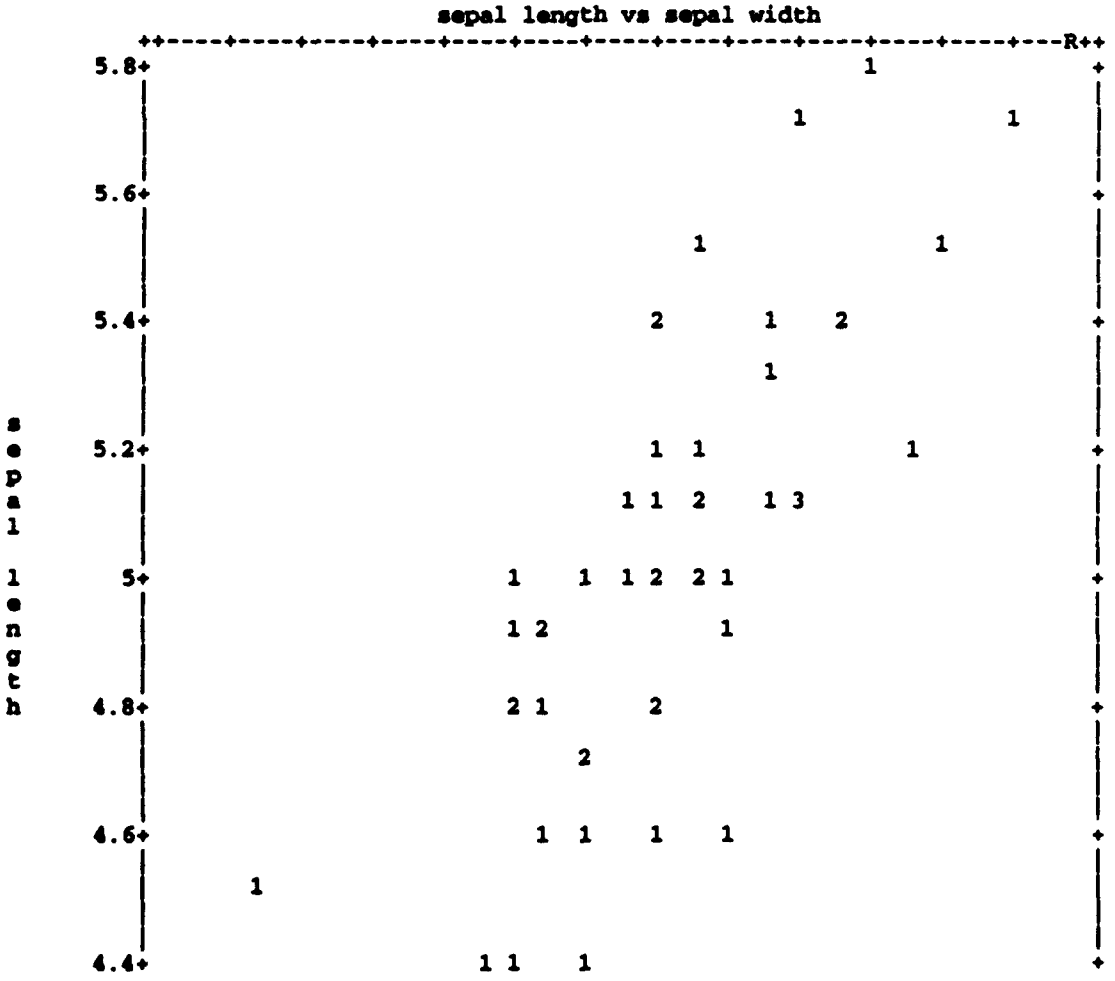
50 unweighted cases accepted.

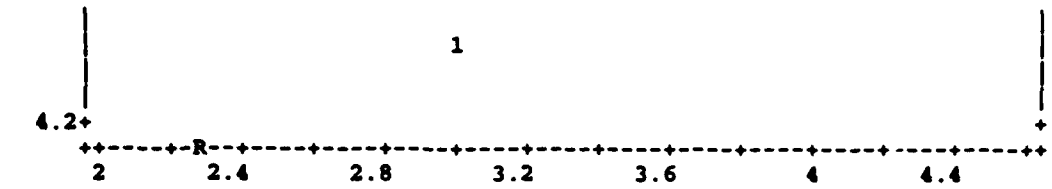
Size of the plots

Horizontal size is 65
Vertical size is 40

Frequencies and symbols used (not applicable for control or overlay plots)

1 - 1	11 - B	21 - L	31 - V
2 - 2	12 - C	22 - M	32 - W
3 - 3	13 - D	23 - N	33 - X
4 - 4	14 - E	24 - O	34 - Y
5 - 5	15 - F	25 - P	35 - Z
6 - 6	16 - G	26 - Q	36 - *
7 - 7	17 - H	27 - R	
8 - 8	18 - I	28 - S	
9 - 9	19 - J	29 - T	
10 - A	20 - K	30 - U	





50 cases plotted. Regression statistics of SL on SW:
Correlation .74255 R Squared .55138 S.E. of Est .23854 Sig. .0000
Intercept(S.E.) 2.63900(.31001) Slope(S.E.) .69049(.08990)

Preceding task required .27 seconds CPU time; .79 seconds elapsed.

```
  39  regression / variables = sl sw
  40           / dependent = sl
  41           / method = enter sw
  42           / scatterplot (*sresid *pred)
  43           / residuals = histogram (sresid)
  44
  45
  46  ************************************************************
  47  * The last subroutine calculates multiple regression for *
  48  * all flower measurements for species 1 only              *
  49  ************************************************************
```

There are 1,497,728 bytes of memory available.

932 bytes of memory required for REGRESSION procedure.
6880 more bytes may be needed for Residuals plots.

*** * * * M U L T I P L E R E G R E S S I O N * * * ***

Listwise Deletion of Missing Data

Equation Number 1 Dependent Variable.. SL sepal length

Block Number 1. Method: Enter SW

Variable(s) Entered on Step Number
1.. SW sepal width

Multiple R	.74255
R Square	.55138
Adjusted R Square	.54203
Standard Error	.23854

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	3.35688	3.35688
Residual	48	2.73132	.05690

F = 58.99373 Signif F = .0000

---------- Variables in the Equation ----------

Variable	B	SE B	Beta	T	Sig T
SW	.690490	.089899	.742547	7.681	.0000
(Constant)	2.639001	.310014		8.513	.0000

End Block Number 1 All requested variables entered.

* * * * M U L T I P L E R E G R E S S I O N * * * *

Equation Number 1 Dependent Variable.. SL sepal length

Residuals Statistics:

	Min	Max	Mean	Std Dev	N
*PRED	4.2271	5.6772	5.0060	.2617	50
*ZPRED	-2.9757	2.5642	.0000	1.0000	50
*SEPRED	.0338	.1069	.0453	.0150	50
*ADJPRED	4.1586	5.6730	5.0050	.2654	50
*RESID	-.5248	.4443	.0000	.2361	50
*ZRESID	-2.1999	1.8625	.0000	.9897	50
*SRESID	-2.2270	1.8821	.0019	1.0099	50
*DRESID	-.5378	.4552	.0010	.2460	50
*SDRESID	-2.3272	1.9352	.0015	1.0270	50
*MAHAL	.0055	8.8551	.9800	1.6399	50
*COOK D	.0000	.2056	.0213	.0357	50
*LEVER	.0001	.1807	.0200	.0335	50

Total Cases = 50

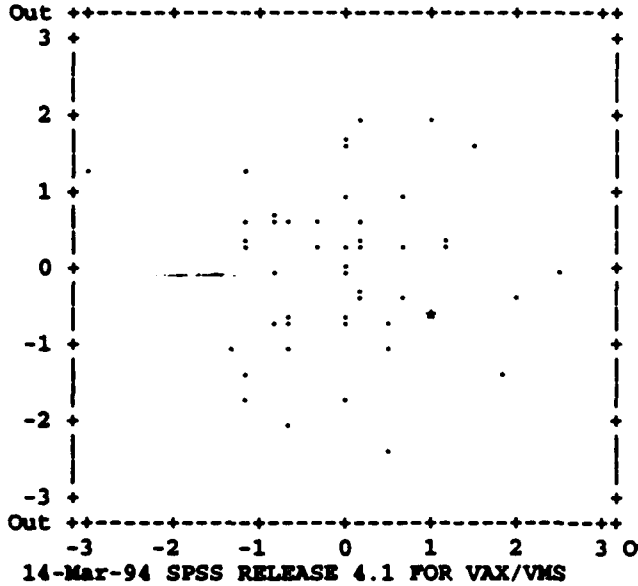
* *

Histogram - Studentized Residual

```
NExp N      (* = 1 Cases,     . : = Normal Curve)
0  .04   Out
0  .08  3.00
0  .20  2.67
0  .45  2.33
2  .91  2.00 :*
3 1.67  1.67 *:*
2 2.74  1.33 **.
2 4.03  1.00 ** .
6 5.31   .67 ****:*
9 6.26   .33 *****:***
4 6.62   .00 ****  .
4 6.26  -.33 **** .
9 5.31  -.67 ****:****
3 4.03 -1.00 ***.
2 2.74 -1.33 **.
2 1.67 -1.67 *:
1  .91 -2.00 :
1  .45 -2.33 *
0  .20 -2.67
```

```
0  .08 -3.00
0  .04   Out
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  12
17:13:31  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


Standardized Scatterplot
Across - *PRED      Down - *SRESID
```




```
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  13
17:13:32  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


Preceding task required .35 seconds CPU time;  .95 seconds elapsed.

  50  regression
  51          / variables = sl sw pl pw
  52          / dependent = sl
  53          / method = stepwise

There are 1,497,856 bytes of memory available.


     1484 bytes of memory required for REGRESSION procedure.
        0 more bytes may be needed for Residuals plots.


 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  14
17:13:32  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

          * * * *  M U L T I P L E  R E G R E S S I O N  * * * *


Listwise Deletion of Missing Data

Equation Number 1    Dependent Variable..   SL   sepal length

Block Number  1.  Method:  Stepwise     Criteria   PIN  .0500   POUT  .1000


Variable(s) Entered on Step Number
   1..     SW         sepal width
```

```
Multiple R            .74255
R Square              .55138
Adjusted R Square     .54203
Standard Error        .23854

Analysis of Variance
                    DF        Sum of Squares      Mean Square
Regression           1              3.35688          3.35688
Residual            48              2.73132           .05690

F =       58.99373       Signif F =  .0000


------------------ Variables in the Equation ------------------

Variable               B         SE B       Beta          T  Sig T

SW               .690490      .089899    .742547      7.681  .0000
(Constant)      2.639001      .310014                 8.513  .0000


------------- Variables not in the Equation -------------

Variable     Beta In  Partial  Min Toler         T  Sig T

PL           .139635  .205156    .968423     1.437  .1573
PW           .111299  .161605    .945827     1.123  .2673


End Block Number   1   PIN =     .050 Limits reached.
 14-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page  15
17:13:32  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


Preceding task required .13 seconds CPU time;  .40 seconds elapsed.

  54  execute

Preceding task required .02 seconds CPU time;  .02 seconds elapsed.

  55  finish

    55 command lines read.
     0 errors detected.
     0 warnings issued.
     2 seconds CPU time.
     4 seconds elapsed time.
       End of job.
```

```
 11-Mar-94   SPSS RELEASE 4.1 FOR VAX/VMS
Page   1
19:42:56    WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

VAX                        WESTERN WASHINGTON UNIVERSITY     License Number 20077
This software is functional through December 31, 1994.

   1  0  file handle iris/name='iris.dat'
   2  0  data list file iris free/ species sl sw pl pw
   3  0  variable labels
   4  0       sl 'sepal length' /
   5  0       sw 'sepal width' /
   6  0       pl 'petal length' /
   7  0       pw 'petal width'/
   8  0
   9  0  set width=80
  10  0
  11  0
  12  0  ***************************************************************************
  13  0  * This calculates ANOVA and multiple ranges for sepal length,      *
  14  0  * lists descriptive statistics, and test homogeneity of variances *
  15  0  ***************************************************************************
  16  oneway sl by species(1,3)
  17           / ranges = duncan
  18           / ranges = snk
  19           / statistics = all
  20
  21  *****************************************************
  22  * This calculates a nonparametric version of ANOVA *
  23  *****************************************************

ONEWAY problem requires 374 bytes of memory.

There are 1,490,016 bytes of memory available.

 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page   2
19:42:57  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

- - - - - - - - - - - - - - - - O N E W A Y - - - - - - - - - - - - - - - - - -

     Variable  SL        sepal length
  By Variable  SPECIES

                                ANALYSIS OF VARIANCE

                                  SUM OF          MEAN            F      F
        SOURCE            D.F.  SQUARES        SQUARES         RATIO  PROB.

BETWEEN GROUPS               2     61.6985       30.8493     106.3508  .0000

WITHIN GROUPS              147     42.6404         .2901

TOTAL                      149    104.3389


                                 STANDARD   STANDARD
GROUP        COUNT       MEAN   DEVIATION      ERROR    95 PCT CONF INT FOR MEAN

Grp 1           50     5.0060       .3525      .0498     4.9058  TO      5.1062
Grp 2           50     5.9160       .5479      .0775     5.7603  TO      6.0717
Grp 3           50     6.5700       .6677      .0944     6.3802  TO      6.7598

TOTAL          150     5.8307       .8368      .0683     5.6957  TO      5.9657
```

```
          FIXED EFFECTS MODEL        .5386       .0440       5.7438  TO       5.9176

          RANDOM EFFECTS MODEL                   .4535       3.8794  TO       7.7819

RANDOM EFFECTS MODEL - ESTIMATE OF BETWEEN COMPONENT VARIANCE          0.6112


GROUP          MINIMUM     MAXIMUM

Grp 1           4.3000      5.8000
Grp 2           4.6000      7.0000
Grp 3           4.9000      7.9000

TOTAL           4.3000      7.9000
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                                  Page   3
19:42:57  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

- - - - - - - - - - - - - - - - O N E W A Y - - - - - - - - - - - - - - - - - -


Tests for Homogeneity of Variances

      Cochrans C = Max. Variance/Sum(Variances) =  .5123, P =  .003 (Approx.)
      Bartlett-Box F =                            9.324 , P =  .000
      Maximum Variance / Minimum Variance         3.588
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                                  Page   4
19:42:57  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

- - - - - - - - - - - - - - - - O N E W A Y - - - - - - - - - - - - - - - - - -


     Variable  SL         sepal length
   By Variable  SPECIES


MULTIPLE RANGE TEST


DUNCAN PROCEDURE
RANGES FOR THE 0.050 LEVEL -

          2.80   2.94

THE RANGES ABOVE ARE TABLE RANGES.
THE VALUE ACTUALLY COMPARED WITH MEAN(J)-MEAN(I) IS..
        0.3808 * RANGE * DSQRT(1/N(I) + 1/N(J))

   (*) DENOTES PAIRS OF GROUPS SIGNIFICANTLY DIFFERENT AT THE 0.050 LEVEL

                          G G G
                          r r r
                          p p p

     Mean       Group     1 2 3

     5.0060     Grp 1
     5.9160     Grp 2     *
     6.5700     Grp 3     * *


  HOMOGENEOUS SUBSETS   (SUBSETS OF GROUPS, WHOSE HIGHEST AND LOWEST MEANS
                         DO NOT DIFFER BY MORE THAN THE SHORTEST
```

SIGNIFICANT RANGE FOR A SUBSET OF THAT SIZE)

SUBSET 1

GROUP Grp 1
MEAN 5.0060
- - - - - - - - - - -

SUBSET 2

GROUP Grp 2
MEAN 5.9160
- - - - - - - - - - -

SUBSET 3

GROUP Grp 3
MEAN 6.5700
- - - - - - - - - - -

- - - - - - - - - - - - - - - O N E W A Y - - - - - - - - - - - - - - - - - -

Variable SL sepal length
By Variable SPECIES

MULTIPLE RANGE TEST

STUDENT-NEWMAN-KEULS PROCEDURE
RANGES FOR THE 0.050 LEVEL -

2.81 3.35

THE RANGES ABOVE ARE TABLE RANGES.
THE VALUE ACTUALLY COMPARED WITH MEAN(J)-MEAN(I) IS..
0.3808 * RANGE * DSQRT(1/N(I) + 1/N(J))

(*) DENOTES PAIRS OF GROUPS SIGNIFICANTLY DIFFERENT AT THE 0.050 LEVEL

		G G G r r r p p p
Mean	Group	1 2 3
5.0060	Grp 1	
5.9160	Grp 2	*
6.5700	Grp 3	* *

HOMOGENEOUS SUBSETS (SUBSETS OF GROUPS, WHOSE HIGHEST AND LOWEST MEANS DO NOT DIFFER BY MORE THAN THE SHORTEST SIGNIFICANT RANGE FOR A SUBSET OF THAT SIZE)

SUBSET 1

```
GROUP       Grp 1
MEAN        5.0060
- - - - - - - - - - -

SUBSET  2

GROUP       Grp 2
MEAN        5.9160
- - - - - - - - - - -
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page   7
19:42:58  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


SUBSET  3

GROUP       Grp 3
MEAN        6.5700
- - - - - - - - - - -
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page   8
19:42:58  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


Preceding task required .48 seconds CPU time;  .76 seconds elapsed.

  24  npar tests k-w = sl by species(1,3)
  25
  26  ************************************************************
  27  * This calculates multiple analysis of variance for        *
  28  * all four flower measurements by species                  *
  29  ************************************************************

There are 1,498,912 bytes of memory available.


***** Workspace allows for 40154 cases for NPAR tests *****

 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page   9
19:42:58  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


- - - - - Kruskal-Wallis 1-Way Anova

    SL        sepal length
  by SPECIES


     Mean Rank    Cases

         30.96       50   SPECIES =    1
         82.26       50   SPECIES =    2
        113.28       50   SPECIES =    3

                    ---

                    150   Total


                                                 Corrected for ties
        Cases     Chi-Square  Significance    Chi-Square  Significance
          150        91.5719         .0000       91.7466         .0000

 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page  10
19:42:58  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


Preceding task required .15 seconds CPU time;  .54 seconds elapsed.

  30  manova sl sw pl pw
```

```
  31          by species(1,3)
  32          / print=homogeneity(all)
  33          / power
  34          / print=signif(efsize)
  35          / cinterval=multivariate(wilks)
  36
  37

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

>Note # 12167
>The last subcommand is not a design specification--A full factorial model is
>generated for this problem.

 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  11
19:42:59  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


* * * * * * A N A L Y S I S   O F   V A R I A N C E * * * * * *

       150 cases accepted.
         0 cases rejected because of out-of-range factor values.
         0 cases rejected because of missing data.
         3 non-empty cells.

         1 design will be processed.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
              CELL NUMBER
                1    2    3
Variable
  SPECIES       1    2    3

Univariate Homogeneity of Variance Tests

Variable .. SL                 sepal length

      Cochrans C(49,3) =                        .51231, P =  .003 (approx.)
      Bartlett-Box F(2,48620) =                9.32439, P =  .000

Variable .. SW                 sepal width

      Cochrans C(49,3) =                        .41509, P =  .215 (approx.)
      Bartlett-Box F(2,48620) =                1.04554, P =  .352

Variable .. PL                 petal length

      Cochrans C(49,3) =                        .53990, P =  .001 (approx.)
      Bartlett-Box F(2,48620) =               27.93313, P =  .000

Variable .. PW                 petal width

      Cochrans C(49,3) =                        .60036, P =  .000 (approx.)
      Bartlett-Box F(2,48620) =               19.62158, P =  .000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Cell Number .. 1

Determinant of Variance-Covariance matrix =          .00000
LOG(Determinant) =                               -13.06736

- - - - - - - - - - -

 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  12
19:42:59  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4
```

```
* * * * * * A N A L Y S I S   O F   V A R I A N C E -- DESIGN   1 * * * * * *

Cell Number .. 2

Determinant of Variance-Covariance matrix =          .00003
LOG(Determinant) =                                 -10.54631

- - - - - - - - - - -

Cell Number .. 3

Determinant of Variance-Covariance matrix =          .00017
LOG(Determinant) =                                  -8.67736

- - - - - - - - - - -


Determinant of pooled Variance-Covariance matrix         .00006
LOG(Determinant) =                                     -9.72392


- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Multivariate test for Homogeneity of Dispersion matrices


Boxs M =                        152.84429
F WITH (20,77566) DF =            7.34218, P =    .000 (Approx.)
Chi-Square with 20 DF =         146.88302, P =    .000 (Approx.)


- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page  13
19:42:59  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


* * * * * * A N A L Y S I S   O F   V A R I A N C E -- DESIGN   1 * * * * * *

EFFECT .. SPECIES
Multivariate Tests of Significance (S = 2, M = 1/2, N = 71 )

Test Name          Value  Approx. F Hypoth. DF   Error DF  Sig. of F

Pillais          1.19643   53.97255       8.00     290.00       .000
Hotellings      32.37084  578.62869       8.00     286.00       .000
Wilks             .02338  199.44317       8.00     288.00       .000
Roys              .96977
Note.. F statistic for WILK'S Lambda is exact.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Multivariate Effect Size and Observed Power at .0500 Level

TEST NAME   Effect Sizent.       P

Pillais             .598    431.780       1.00
Hotellings          .942   4629.030       1.00
Wilks               .847   1595.545       1.00

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
EFFECT .. SPECIES (Cont.)
Univariate F-tests with (2,147) D. F.

Variable   Hypoth. SS   Error SS Hypoth. MS   Error MS          F  Sig. of F

SL           61.69853   42.64040   30.84927     .29007  106.35084       .000
SW           11.34493   16.96200    5.67247     .11539   49.16004       .000
```

```
PL          437.71000   27.64340  218.85500      .18805 1163.81071        .000
PW           80.41333    6.15660   40.20667      .04188  960.00715        .000

Variable   ETA Square   Noncent.      Power

SL             .59133  212.70167    1.00000
SW             .40078   98.32008    1.00000
PL             .94060 2327.62142    1.00000
PW             .92888 1920.01429    1.00000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Estimates for SL
--- Individual multivariate .9500 WILK confidence intervals
--- two-tailed observed power taken at .0500 level
11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  14
19:43:00  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


* * * * * * A N A L Y S I S   O F   V A R I A N C E -- DESIGN   1 * * * * * *

Estimates for SL  (Cont.)
SPECIES

 Parameter      Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

       2   -.82466667      .06219  -13.26041      .00000   -1.02075    -.62859
       3   .085333333      .06219    1.37214      .17211    -.11075     .28141

 Parameter   Noncent.  Power

       2    175.83860      1.000
       3      1.88276       .274

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Estimates for SW
--- Individual multivariate .9500 WILK confidence intervals
--- two-tailed observed power taken at .0500 level
SPECIES

 Parameter      Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

       2   .370666667      .03922    9.45005      .00000     .24700     .49434
       3   -.28733333      .03922   -7.32549      .00000    -.41100    -.16366

 Parameter   Noncent.  Power

       2     89.30353      1.000
       3     53.66283      1.000

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Estimates for PL
--- Individual multivariate .9500 WILK confidence intervals
--- two-tailed observed power taken at .0500 level
SPECIES

 Parameter      Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper

       2   -2.3000000      .05007  -45.93264      .00000   -2.45788   -2.14212
       3   .510000000      .05007   10.18506      .00000     .35212     .66788

 Parameter   Noncent.  Power

       2   2109.80740      1.000
       3    103.73552      1.000

11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  15
19:43:00  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4
```

```
* * * * * * A N A L Y S I S   O F   V A R I A N C E -- DESIGN   1 * * * * * *

 Estimates for PW
 --- Individual multivariate .9500 WILK confidence intervals
 --- two-tailed observed power taken at .0500 level
 SPECIES

  Parameter      Coeff.  Std. Err.     t-Value     Sig. t Lower -95%  CL- Upper

        2   -.95333333     .02363   -40.34257     .00000   -1.02784     -.87883
        3    .126666667    .02363     5.36020     .00000     .05216      .20117

  Parameter   Noncent.  Power

        2   1627.52331     1.000
        3     28.73177     1.000

 - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

6768 bytes of memory are needed for MANOVA execution.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page  16
19:43:00  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


Preceding task required .68 seconds CPU time;  1.52 seconds elapsed.

  38  execute

Preceding task required .02 seconds CPU time;  .02 seconds elapsed.

  39  finish

    39 command lines read.
     0 errors detected.
     0 warnings issued.
     2 seconds CPU time.
     4 seconds elapsed time.
       End of job.
```

```
 11-Mar-94    SPSS RELEASE 4.1 FOR VAX/VMS
Page    1
16:16:03     WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

VAX                        WESTERN WASHINGTON UNIVERSITY      License Number 20077
This software is functional through December 31, 1994.

   1  0  file handle iris/name='iris.dat'
   2  0  data list file iris free/ species sl sw pl pw
   3  0  variable labels
   4  0       sl 'sepal length' /
   5  0       sw 'sepal width' /
   6  0       pl 'petal length' /
   7  0       pw 'petal width'/
   8  0
   9  0  set width=80
  10  0
  11  0  ************************************************************
  12  0  * This clusters the iris data using sq euclidean distance *
  13  0  * with the centroid clustering method                      *
  14  0  ************************************************************
  15  cluster sl sw pl pw
  16  / plot = dendrogram
  17  / method = centroid
  18  / measure = seuclid
  19
  20  ******************************************************
  21  * This clusters the iris data using cosine distance *
  22  * with the average linkage clustering method        *
  23  ******************************************************

There are 1,496,960 bytes of memory available.


CLUSTER requires 50144 bytes of workspace for execution.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   2
16:16:05  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S * * * * * *


 Data Information

        150 unweighted cases accepted.
          0 cases rejected because of missing value.

 Squared Euclidean measure used.

 1 Agglomeration method specified.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page   3
16:16:06  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S * * * * * *


  Agglomeration Schedule using Centroid Method

                Clusters  Combined                 Stage Cluster 1st Appears    Next
  Stage      Cluster 1  Cluster 2    Coefficient   Cluster 1   Cluster 2      Stage

      1            102        143       .000000           0           0         58
      2            129        133       .010000           0           0         73
      3             11         49       .010000           0           0         63
      4              8         40       .010000           0           0         22
      5             10         35       .010000           0           0         24
      6              1         18       .010000           0           0         21
```

7	128	139	.020000	0	0	51
8	117	138	.020000	0	0	45
9	97	100	.020000	0	0	28
10	58	94	.020000	0	0	93
11	83	93	.020000	0	0	52
12	64	92	.020000	0	0	37
13	81	82	.020000	0	0	36
14	4	48	.020000	0	0	30
15	20	47	.020000	0	0	29
16	2	46	.020000	0	0	23
17	9	39	.020000	0	0	53
18	5	38	.020000	0	0	44
19	30	31	.020000	0	0	39
20	28	29	.020000	0	0	21
21	1	28	.017500	6	20	22
22	1	8	.021875	21	4	31
23	2	13	.025000	16	0	24
24	2	10	.026944	23	5	38
25	113	140	.030000	0	0	89
26	124	127	.030000	0	0	75
27	89	96	.030000	0	0	28
28	89	97	.027500	27	9	46
29	20	22	.035000	15	0	63
30	3	4	.035000	0	14	54
31	1	50	.038889	22	0	35
32	75	98	.040000	0	0	91
33	54	90	.040000	0	0	74
34	24	27	.040000	0	0	43
35	1	41	.042857	31	0	44
36	70	81	.045000	0	13	74
37	64	79	.045000	12	0	65
38	2	26	.047200	24	0	39
39	2	30	.049444	38	19	68
40	111	148	.050000	0	0	103
41	121	144	.050000	0	0	55
42	66	87	.050000	0	0	66
43	24	44	.050000	34	0	90
44	1	5	.051563	35	18	78
45	104	117	.055000	0	8	96

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Agglomeration Schedule using Centroid Method (CONT.)

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
46	89	95	.059375	28	0	81
47	137	149	.060000	0	0	77
48	142	146	.060000	0	0	89
49	141	145	.060000	0	0	55
50	55	59	.060000	0	0	66
51	71	128	.065000	0	7	111
52	68	83	.065000	0	11	81
53	9	43	.065000	17	0	62
54	3	7	.065556	30	0	68
55	121	141	.067500	41	49	72
56	108	131	.070000	0	0	105
57	106	123	.070000	0	0	109
58	102	114	.070000	1	0	67
59	69	88	.070000	0	0	130
60	52	57	.070000	0	0	97
61	51	53	.070000	0	0	92
62	9	14	.075556	53	0	114

63	11	20	.078056	3	29	78
64	21	32	.080000	0	0	84
65	64	74	.083333	37	0	108
66	55	66	.087500	50	42	70
67	102	122	.087778	58	0	110
68	2	3	.088438	39	54	88
69	76	77	.090000	0	0	70
70	55	76	.074375	66	69	92
71	12	25	.090000	0	0	101
72	121	125	.091875	55	0	121
73	105	129	.092500	0	2	98
74	54	70	.096667	33	36	112
75	124	147	.097500	26	0	87
76	56	91	.100000	0	0	94
77	116	137	.105000	0	47	122
78	1	11	.105000	44	63	85
79	84	134	.110000	0	0	86
80	6	19	.110000	0	0	106
81	68	89	.112222	52	46	94
82	126	130	.120000	0	0	104
83	33	34	.120000	0	0	95
84	21	37	.120000	64	0	85
85	1	21	.126533	78	84	90
86	73	84	.127500	0	79	87
87	73	124	.112222	86	75	124
88	2	36	.128681	68	0	101
89	113	142	.132500	25	48	103

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Agglomeration Schedule using Centroid Method (CONT.)

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
90	1	24	.136389	85	43	120
91	72	75	.140000	0	32	108
92	51	55	.144722	61	70	99
93	58	99	.145000	10	0	119
94	56	68	.145781	76	81	100
95	17	33	.150000	0	83	106
96	104	112	.154444	45	0	98
97	52	86	.157500	60	0	111
98	104	105	.157986	96	73	117
99	51	78	.158906	92	0	128
100	56	62	.166300	94	0	118
101	2	12	.169068	88	71	114
102	118	132	.170000	0	0	140
103	111	113	.170625	40	89	117
104	103	126	.180000	0	82	105
105	103	108	.170833	104	56	131
106	6	17	.184167	80	95	115
107	65	80	.200000	0	0	112
108	64	72	.212431	65	91	123
109	106	119	.217500	57	0	139
110	102	115	.229375	67	0	126
111	52	71	.234444	97	51	123
112	54	65	.234800	74	107	113
113	54	60	.234082	112	0	118
114	2	9	.238656	101	62	125
115	6	15	.238800	106	0	116
116	6	16	.232500	115	0	142
117	104	111	.239762	98	103	121
118	54	56	.250975	113	100	132

119	58	61	.251111	93	0	147
120	1	45	.268934	90	0	125
121	104	121	.269368	117	72	122
122	104	116	.229259	121	77	133
123	52	64	.271310	111	108	128
124	73	120	.306944	87	0	126
125	1	2	.360690	120	114	129
126	73	102	.372882	124	110	134
127	67	107	.380000	0	0	135
128	51	52	.385779	99	123	134
129	1	23	.407008	125	0	142
130	63	69	.427500	0	59	137
131	103	136	.430000	105	0	139
132	54	85	.457008	118	0	137
133	101	104	.497483	0	122	138

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Agglomeration Schedule using Centroid Method (CONT.)

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
134	51	73	.515499	128	126	143
135	67	150	.525000	127	0	145
136	109	135	.570000	0	0	138
137	54	63	.595117	132	130	143
138	101	109	.678967	133	136	144
139	103	106	.693056	131	109	141
140	110	118	.762500	0	102	141
141	103	110	.763210	139	140	144
142	1	6	.836225	129	116	146
143	51	54	.971792	134	137	145
144	101	103	1.475938	138	141	148
145	51	67	1.498889	143	135	147
146	1	42	1.621804	142	0	149
147	51	58	2.847609	145	119	148
148	51	101	3.306975	147	144	149
149	1	51	15.786713	146	148	0

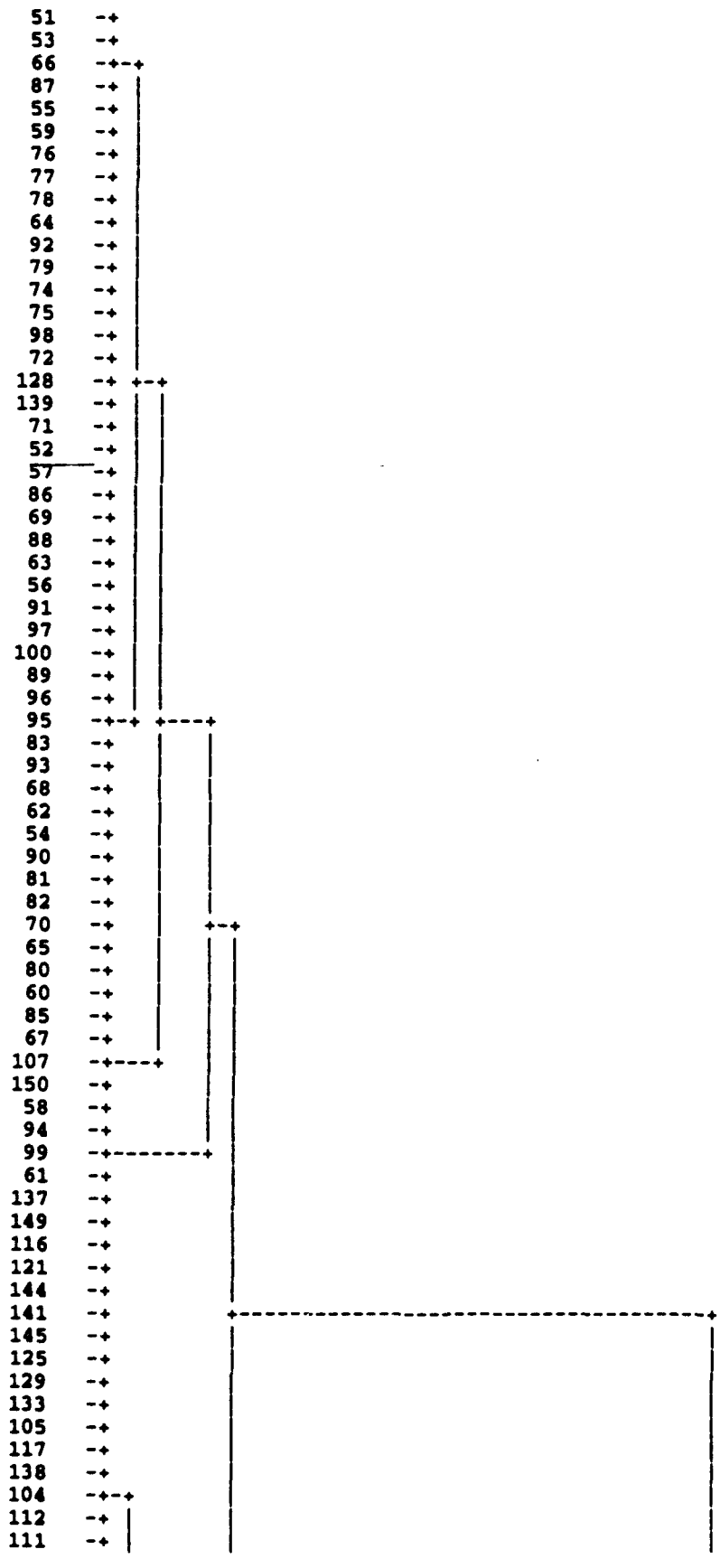
* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

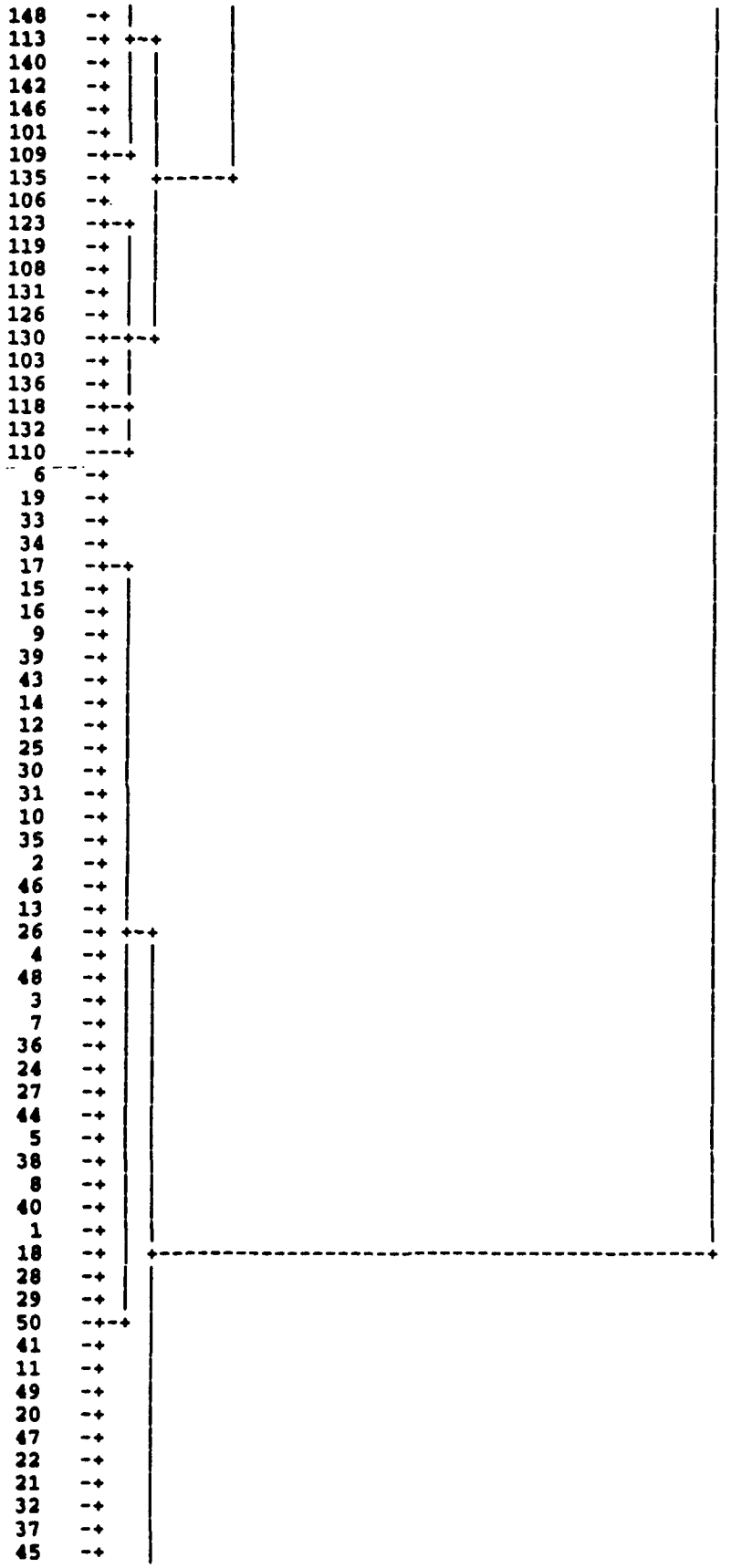
Dendrogram using Centroid Method

Rescaled Distance Cluster Combine

```
  C A S E    0         5        10        15        20        25
  Label  Seq +---------+---------+---------+---------+---------+

        102  -+
        143  -+
        114  -+
        122  -+
        115  -+
        124  -+
        127  -+
        147  -+
         84  -+
        134  -+
         73  -+
        120  -+
```


51
53
66
87
55
59
76
77
78
64
92
79
74
75
98
72
128
139
71
52
57
86
69
88
63
56
91
97
100
89
96
95
83
93
68
62
54
90
81
82
70
65
80
60
85
67
107
150
58
94
99
61
137
149
116
121
144
141
145
125
129
133
105
117
138
104
112
111


148
113
140
142
146
101
109
135
106
123
119
108
131
126
130
103
136
118
132
110
6
19
33
34
17
15
16
9
39
43
14
12
25
30
31
10
35
2
46
13
26
4
48
3
7
36
24
27
44
5
38
8
40
1
18
28
29
50
41
11
49
20
47
22
21
32
37
45

```
          23   -+   |
          42   -----+
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page  11
16:16:09  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

Preceding task required 1.95 seconds CPU time;  4.61 seconds elapsed.

  24  cluster sl sw pl pw
  25  / plot = dendrogram
  26  / method = baverage
  27  / measure = cosine
  28
  29

There are 1,497,376 bytes of memory available.

CLUSTER requires 50144 bytes of workspace for execution.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page  12
16:16:09  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S * * * * * *

 Data Information

        150 unweighted cases accepted.
          0 cases rejected because of missing value.

 Cosine measure used.

 1 Agglomeration method specified.

- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                              Page  13
16:16:10  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S * * * * * *

  Agglomeration Schedule using Average Linkage (Between Groups)
```

Stage	Clusters Combined Cluster 1	Cluster 2	Coefficient	Stage Cluster 1st Appears Cluster 1	Cluster 2	Next Stage
1	102	143	1.000000	0	0	24
2	103	112	.999998	0	0	87
3	66	83	.999997	0	0	17
4	1	11	.999995	0	0	10
5	51	75	.999992	0	0	17
6	114	133	.999989	0	0	37
7	36	37	.999989	0	0	121
8	59	81	.999988	0	0	13
9	136	147	.999987	0	0	87
10	1	3	.999986	4	0	27
11	8	39	.999983	0	0	23
12	28	40	.999977	0	0	23
13	59	70	.999977	8	0	25
14	92	95	.999976	0	0	45
15	128	139	.999975	0	0	41
16	118	138	.999973	0	0	77
17	51	66	.999967	5	3	21
18	21	35	.999967	0	0	61
19	29	50	.999959	0	0	74
20	61	77	.999957	0	0	96
21	51	94	.999956	17	0	36
22	109	123	.999952	0	0	65

23	8	28	.999951	11	12	39
24	102	105	.999948	1	0	38
25	59	93	.999946	13	0	59
26	7	20	.999945	0	0	62
27	1	49	.999945	10	0	64
28	127	140	.999941	0	0	46
29	84	117	.999940	0	0	77
30	4	9	.999939	0	0	53
31	57	60	.999932	0	0	95
32	120	131	.999930	0	0	85
33	10	13	.999929	0	0	61
34	87	98	.999923	0	0	78
35	5	43	.999919	0	0	66
36	51	72	.999916	21	0	73
37	114	129	.999915	6	0	71
38	102	144	.999915	24	0	71
39	8	19	.999914	23	0	54
40	137	149	.999910	0	0	115
41	111	128	.999910	0	15	103
42	56	64	.999906	0	0	60
43	97	100	.999901	0	0	72
44	110	116	.999893	0	0	58
45	79	92	.999891	0	14	98

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Agglomeration Schedule using Average Linkage (Between Groups) (CONT.)

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
46	127	148	.999890	28	0	63
47	52	62	.999886	0	0	72
48	126	134	.999884	0	0	101
49	32	46	.999882	0	0	108
50	18	41	.999878	0	0	93
51	53	90	.999865	0	0	69
52	6	22	.999861	0	0	62
53	4	30	.999861	30	0	67
54	8	48	.999860	39	0	64
55	121	141	.999855	0	0	90
56	54	55	.999853	0	0	86
57	89	96	.999851	0	0	92
58	110	145	.999849	44	0	68
59	59	82	.999822	25	0	89
60	56	91	.999822	42	0	84
61	10	21	.999822	33	18	76
62	6	7	.999801	52	26	100
63	113	127	.999799	0	46	70
64	1	8	.999799	27	54	74
65	106	109	.999799	0	22	79
66	5	38	.999795	35	0	82
67	4	31	.999791	53	0	99
68	110	122	.999785	58	0	90
69	53	76	.999783	51	0	88
70	113	124	.999779	63	0	107
71	102	114	.999772	38	37	112
72	52	97	.999766	47	43	92
73	51	58	.999758	36	0	78
74	1	29	.999750	64	19	93
75	142	146	.999743	0	0	120
76	2	10	.999740	0	61	94
77	84	118	.999739	29	16	83
78	51	87	.999736	73	34	89

79	106	108	.999722	65	0	85
80	73	130	.999718	0	0	101
81	71	85	.999711	0	0	140
82	5	47	.999711	66	0	100
83	84	104	.999696	77	0	125
84	56	132	.999695	60	0	98
85	106	120	.999678	79	32	126
86	54	78	.999670	56	0	88
87	103	136	.999645	2	9	129
88	53	54	.999637	69	86	116
89	51	59	.999637	78	59	104

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Agglomeration Schedule using Average Linkage (Between Groups) (CONT.)

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
90	110	121	.999636	68	55	112
91	16	34	.999632	0	0	109
92	52	89	.999611	72	57	95
93	1	18	.999588	74	50	105
94	2	26	.999566	76	0	108
95	52	57	.999560	92	31	114
96	61	88	.999544	20	0	113
97	24	27	.999542	0	0	110
98	56	79	.999534	84	45	106
99	4	12	.999529	67	0	122
100	5	6	.999505	82	62	118
101	73	126	.999501	80	48	125
102	101	107	.999477	0	0	115
103	111	125	.999475	41	0	107
104	51	68	.999411	89	0	127
105	1	14	.999396	93	0	118
106	56	74	.999393	98	0	116
107	111	113	.999332	103	70	120
108	2	32	.999310	94	49	122
109	16	23	.999296	91	0	119
110	24	44	.999281	97	0	137
111	65	99	.999269	0	0	144
112	102	110	.999209	71	90	128
113	61	63	.999205	96	0	131
114	52	86	.999160	95	0	132
115	101	137	.999114	102	40	124
116	53	56	.999110	88	106	132
117	25	45	.999105	0	0	142
118	1	5	.999099	105	100	134
119	16	17	.999096	109	0	130
120	111	142	.998965	107	75	129
121	15	36	.998942	0	7	139
122	2	4	.998937	108	99	134
123	67	150	.998915	0	0	148
124	101	115	.998798	115	0	128
125	73	84	.998775	101	83	135
126	106	119	.998763	85	0	133
127	51	80	.998715	104	0	136
128	101	102	.998683	124	112	138
129	103	111	.998669	87	120	135
130	16	33	.998666	119	0	143
131	61	69	.998607	113	0	141
132	52	53	.998501	114	116	136
133	106	135	.998420	126	0	145

16:16:11 WESTERN WASHINGTON UNIVERSITY on NESSIE VMS V5.4

* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Agglomeration Schedule using Average Linkage (Between Groups) (CONT.)

Stage	Clusters Combined Cluster 1	Cluster 2	Coefficient	Stage Cluster 1st Appears Cluster 1	Cluster 2	Next Stage
134	1	2	.998373	118	122	137
135	73	103	.998306	125	129	138
136	51	52	.998198	127	132	141
137	1	24	.998064	134	110	139
138	73	101	.997631	135	128	140
139	1	15	.997577	137	121	142
140	71	73	.997149	81	138	145
141	51	61	.996819	136	131	144
142	1	25	.996550	139	117	143
143	1	16	.996356	142	130	147
144	51	65	.995959	141	111	146
145	71	106	.995934	140	133	146
146	51	71	.993265	144	145	148
147	1	42	.990918	143	0	149
148	51	67	.988577	146	123	149
149	1	51	.904194	147	148	0

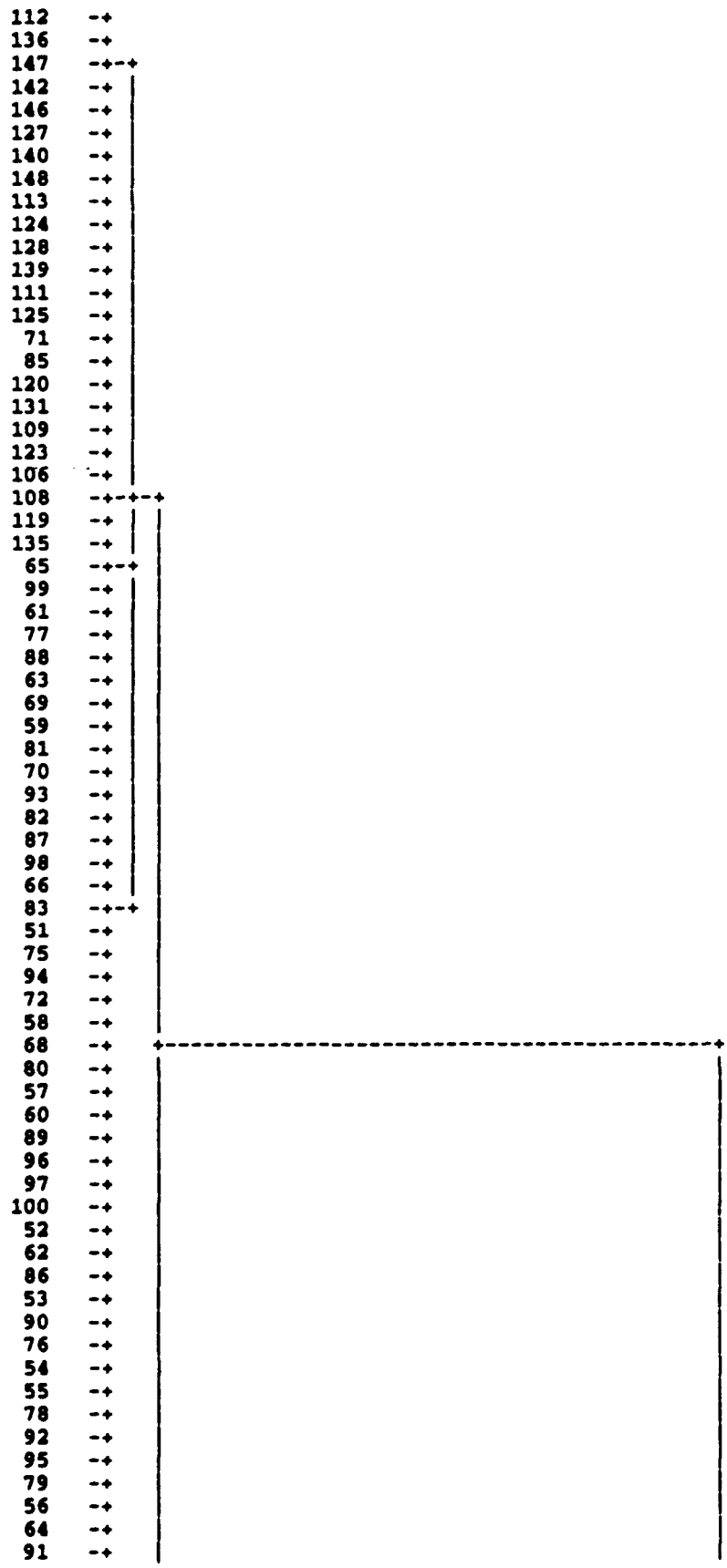
11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS Page 17
16:16:12 WESTERN WASHINGTON UNIVERSITY on NESSIE VMS V5.4

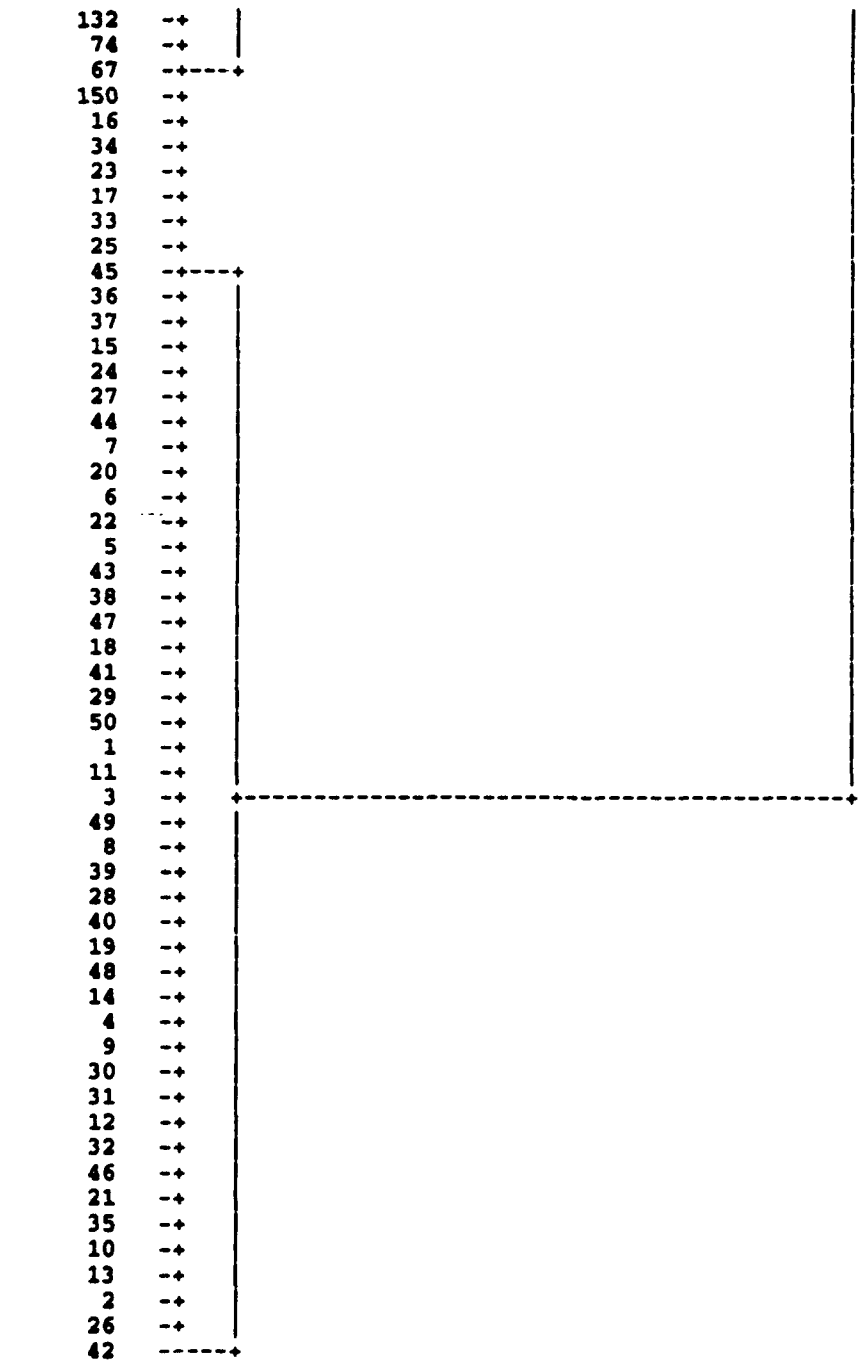
* * * * * * H I E R A R C H I C A L C L U S T E R A N A L Y S I S * * * * * *

Dendrogram using Average Linkage (Between Groups)

Rescaled Distance Cluster Combine

CASE Label	Seq	0 5 10 15 20 25
	102	-+
	143	-+
	105	-+
	144	-+
	114	-+
	133	-+
	129	-+
	121	-+
	141	-+
	110	-+
	116	-+
	145	-+
	122	-+
	137	-+
	149	-+
	101	-+
	107	-+
	115	-+
	118	-+
	138	-+
	84	-+
	117	-+
	104	-+
	126	-+
	134	-+
	73	-+
	130	-+
	103	-+


112
136
147
142
146
127
140
148
113
124
128
139
111
125
71
85
120
131
109
123
106
108
119
135
65
99
61
77
88
63
69
59
81
70
93
82
87
98
66
83
51
75
94
72
58
68
80
57
60
89
96
97
100
52
62
86
53
90
76
54
55
78
92
95
79
56
64
91



Preceding task required 1.50 seconds CPU time; 4.56 seconds elapsed.

30 execute

Preceding task required .01 seconds CPU time; .07 seconds elapsed.

31 finish

31 command lines read.
0 errors detected.

```
 11-Mar-94   SPSS RELEASE 4.1 FOR VAX/VMS
Page   1
19:02:44    WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

VAX                        WESTERN WASHINGTON UNIVERSITY     License Number 20077
This software is functional through December 31, 1994.

   1  0  file handle iris/name='iris.dat'
   2  0  data list file iris free/ species sl sw pl pw
   3  0  variable labels
   4  0       sl 'sepal length' /
   5  0       sw 'sepal width' /
   6  0       pl 'petal length' /
   7  0       pw 'petal width'/
   8  0
   9  0  set width=80
  10  0
  11  0  ************************************************************************
  12  0  * This computes standard PCA and plots the first two components *
  13  0  ************************************************************************
  14  factor var=sl sw pl pw
  15  / criteria = factors(4)
  16  / plot = rotation(1,2)
  17

There are 1,496,768 bytes of memory available.


This FACTOR analysis requires 2864 (       2.8K) bytes of memory.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                                  Page   2
19:02:45  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

- - - - - - - - - - -  F A C T O R  A N A L Y S I S  - - - - - - - - - - -


ANALYSIS NUMBER 1   LISTWISE DELETION OF CASES WITH MISSING VALUES


EXTRACTION  1  FOR ANALYSIS  1, PRINCIPAL-COMPONENTS ANALYSIS (PC)


INITIAL STATISTICS:

VARIABLE      COMMUNALITY  *  FACTOR   EIGENVALUE    PCT OF VAR    CUM PCT
                           *
SL                1.00000  *      1       2.89686         72.4       72.4
SW                1.00000  *      2        .91513         22.9       95.3
PL                1.00000  *      3        .16506          4.1       99.4
PW                1.00000  *      4        .02295           .6      100.0

   PC EXTRACTED   4 FACTORS.


FACTOR MATRIX:

               FACTOR  1      FACTOR  2      FACTOR  3      FACTOR  4

SL               .88055         .36852        -.29596        -.03496
SW              -.46244         .88030         .10453         .01740
PL               .98995         .02286         .07075         .12033
```

PW	.96314	.06224	.24805	-.08336

FINAL STATISTICS:

VARIABLE	COMMUNALITY	*	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
		*				
SL	1.00000	*	1	2.89686	72.4	72.4
SW	1.00000	*	2	.91513	22.9	95.3
PL	1.00000	*	3	.16506	4.1	99.4
PW	1.00000	*	4	.02295	.6	100.0

- - - - - - - - - - - F A C T O R A N A L Y S I S - - - - - - - - - - -

VARIMAX ROTATION 1 FOR EXTRACTION 1 IN ANALYSIS 1 - KAISER NORMALIZATION.

VARIMAX CONVERGED IN 6 ITERATIONS.

ROTATED FACTOR MATRIX:

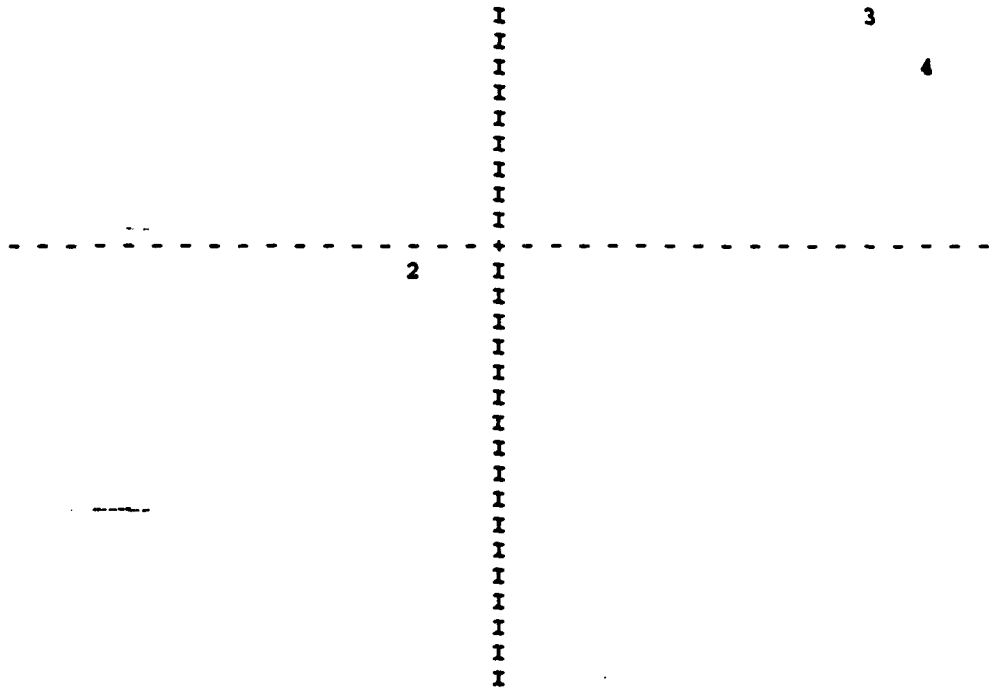
	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
SL	.50947	.86048	.00373	-.00180
SW	-.17756	-.03202	.98356	-.00810
PL	.79103	.52675	-.27417	.14708
PW	.89077	.40366	-.19882	-.06369

FACTOR TRANSFORMATION MATRIX:

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
FACTOR 1	.74969	.58088	-.31567	.02983
FACTOR 2	.11471	.35633	.92725	-.00917
FACTOR 3	.65177	-.73079	.19986	-.03457
FACTOR 4	.00122	-.03936	.02486	.99891

- - - - - - - - - - - F A C T O R A N A L Y S I S - - - - - - - - - - -

HORIZONTAL FACTOR 1 VERTICAL FACTOR 2



SYMBOL	VARIABLE	COORDINATES	SYMBOL	VARIABLE	COORDINATES
1	SL	(.50947, .86048)	2	SW	(-.17756, -.03202)
3	PL	(.79103, .52675)	4	PW	(.89077, .40366)

Preceding task required .48 seconds CPU time; 1.10 seconds elapsed.

18 execute

Preceding task required .02 seconds CPU time; .02 seconds elapsed.

19 finish

19 command lines read.
0 errors detected.
0 warnings issued.
1 seconds CPU time.
2 seconds elapsed time.
End of job.

```
 11-Mar-94    SPSS RELEASE 4.1 FOR VAX/VMS
Page    1
19:20:07     WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4

VAX                        WESTERN WASHINGTON UNIVERSITY     License Number 20077
This software is functional through December 31, 1994.

   1   0  file handle iris/name='iris.dat'
   2   0  data list file iris free/ species sl sw pl pw
   3   0  variable labels
   4   0       sl 'sepal length' /
   5   0       sw 'sepal width' /
   6   0       pl 'petal length' /
   7   0       pw 'petal width'/
   8   0
   9   0  set width=80
  10   0
  11   0  ************************************************************
  12   0  * This procedure does a discriminant analysis on the iris data  *
  13   0  ************************************************************
  14   discriminant groups=species(1,3)
  15   / variables sl sw pl pw
  16   / method = wilks
  17   / statistics = all
  18   / plot
  19

There are 1,496,640 bytes of memory available.


SINCE ANALYSIS= WAS OMITTED FOR THE FIRST ANALYSIS ALL VARIABLES
ON THE VARIABLES= LIST WILL BE ENTERED AT LEVEL 1.


This DISCRIMINANT analysis requires 1160 bytes of memory.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                          Page    2
19:20:08  WESTERN WASHINGTON UNIVERSITY  on NESSIE          VMS V5.4


- - - - - - - -  D I S C R I M I N A N T   A N A L Y S I S  - - - - - - - -

ON GROUPS DEFINED BY SPECIES


           150 (UNWEIGHTED) CASES WERE PROCESSED.
             0 OF THESE WERE EXCLUDED FROM THE ANALYSIS.
           150 (UNWEIGHTED) CASES WILL BE USED IN THE ANALYSIS.


NUMBER OF CASES BY GROUP

                  NUMBER OF CASES
  SPECIES    UNWEIGHTED     WEIGHTED  LABEL

         1            50         50.0
         2            50         50.0
         3            50         50.0

     TOTAL           150        150.0


GROUP MEANS
```

SPECIES	SL	SW	PL	PW
1	5.00600	3.42800	1.46200	0.24600
2	5.91600	2.77000	4.27200	1.32600
3	6.57000	2.97400	5.55200	2.02600
TOTAL	5.83067	3.05733	3.76200	1.19933

GROUP STANDARD DEVIATIONS

SPECIES	SL	SW	PL	PW
1	0.35249	0.37906	0.17366	0.10539
2	0.54786	0.31380	0.47896	0.19775
3	0.66769	0.32250	0.55189	0.27465
TOTAL	0.83682	0.43587	1.76725	0.76224

POOLED WITHIN-GROUPS COVARIANCE MATRIX WITH 147 DEGREES OF FREEDOM

	SL	SW	PL	PW
SL	0.2900707			
SW	0.9099728E-01	0.1153878		
PL	0.1715088	0.5625034E-01	0.1880503	
PW	0.3860136E-01	0.3271020E-01	0.4296735E-01	0.4188163E-01

POOLED WITHIN-GROUPS CORRELATION MATRIX

	SL	SW	PL	PW
SL	1.00000			
SW	0.49739	1.00000		
PL	0.73434	0.38186	1.00000	
PW	0.35022	0.47053	0.48416	1.00000

CORRELATIONS WHICH CANNOT BE COMPUTED ARE PRINTED AS 99.0.

WILKS' LAMBDA (U-STATISTIC) AND UNIVARIATE F-RATIO
WITH 2 AND 147 DEGREES OF FREEDOM

VARIABLE	WILKS' LAMBDA	F	SIGNIFICANCE
SL	0.40867	106.4	0.0000
SW	0.59922	49.16	0.0000
PL	0.05940	1164.	0.0000
PW	0.07112	960.0	0.0000

COVARIANCE MATRIX FOR GROUP 1,

	SL	SW	PL	PW
SL	0.1242490			
SW	0.9921633E-01	0.1436898		
PL	0.1635510E-01	0.1169796E-01	0.3015918E-01	
PW	0.1033061E-01	0.9297959E-02	0.6069388E-02	0.1110612E-01

COVARIANCE MATRIX FOR GROUP 2,

	SL	SW	PL	PW
SL	0.3001469			
SW	0.8048980E-01	0.9846939E-01		
PL	0.1865796	0.8567347E-01	0.2294041	
PW	0.5222857E-01	0.4120408E-01	0.7400816E-01	0.3910612E-01

COVARIANCE MATRIX FOR GROUP 3,

	SL	SW	PL	PW
SL	0.4458163			
SW	0.9328571E-01	0.1040041		
PL	0.3115918	0.7137959E-01	0.3045878	
PW	0.5324490E-01	0.4762857E-01	0.4882449E-01	0.7543265E-01

TOTAL COVARIANCE MATRIX WITH 149 DEGREES OF FREEDOM

	SL	SW	PL	PW
SL	0.7002613			
SW	-0.4170291E-01	0.1899794		
PL	1.264395	-0.3298201	3.123177	
PW	0.5106246	-0.1216394	1.296417	0.5810063

- - - - - - - - - D I S C R I M I N A N T A N A L Y S I S - - - - - - - - -

ON GROUPS DEFINED BY SPECIES

ANALYSIS NUMBER 1

STEPWISE VARIABLE SELECTION

SELECTION RULE: MINIMIZE WILKS' LAMBDA
MAXIMUM NUMBER OF STEPS..... 8
MINIMUM TOLERANCE LEVEL..... 0.00100
MINIMUM F TO ENTER..... 1.0000
MAXIMUM F TO REMOVE..... 1.0000

CANONICAL DISCRIMINANT FUNCTIONS

MAXIMUM NUMBER OF FUNCTIONS..... 2
MINIMUM CUMULATIVE PERCENT OF VARIANCE... 100.00
MAXIMUM SIGNIFICANCE OF WILKS' LAMBDA.... 1.0000

PRIOR PROBABILITY FOR EACH GROUP IS 0.33333

---------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 0 ----------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	WILKS' LAMBDA
SL	1.0000000	1.0000000	106.35	0.40867
SW	1.0000000	1.0000000	49.160	0.59922
PL	1.0000000	1.0000000	1163.8	0.05940
PW	1.0000000	1.0000000	960.01	0.07112

* *

AT STEP 1, PL WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF FREEDOM		SIGNIF.	BETWEEN GROUPS
WILKS' LAMBDA	0.05940	1	2 147.0		
EQUIVALENT F	1163.81		2 147.0	0.0000	

---------- VARIABLES IN THE ANALYSIS AFTER STEP 1 ----------

VARIABLE	TOLERANCE	F TO REMOVE	WILKS' LAMBDA
PL	1.0000000	1163.8	

---------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 1 ----------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	WILKS' LAMBDA
SL	0.4607440	0.4607440	31.295	0.04158
SW	0.8541802	0.8541802	43.475	0.03723
PW	0.7655883	0.7655883	25.927	0.04383

F STATISTICS AND SIGNIFICANCES BETWEEN PAIRS OF GROUPS AFTER STEP 1
EACH F STATISTIC HAS 1 AND 147.0 DEGREES OF FREEDOM.

GROUP	1	2
GROUP		

```
       2                1049.7
                        0.0000

       3                2223.9          217.81
                        0.0000          0.0000
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                                  Page   7
19:20:09  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *


AT STEP   2, SW        WAS INCLUDED IN THE ANALYSIS.

                                    DEGREES OF FREEDOM  SIGNIF.   BETWEEN GROUPS
WILKS' LAMBDA          0.03723       2     2      147.0
EQUIVALENT F           305.332             4      292.0  0.0000


---------------- VARIABLES IN THE ANALYSIS AFTER STEP   2 ----------------

VARIABLE  TOLERANCE  F TO REMOVE   WILKS' LAMBDA

SW        0.8541802    43.475          0.05940
PL        0.8541802    1101.9          0.59922

---------------- VARIABLES NOT IN THE ANALYSIS AFTER STEP   2 ----------------

                     MINIMUM
VARIABLE  TOLERANCE  TOLERANCE  F TO ENTER    WILKS' LAMBDA

SL        0.4056307  0.4056307    11.214          0.03224
PW        0.6700620  0.6700620    35.822          0.02492


F STATISTICS AND SIGNIFICANCES BETWEEN PAIRS OF GROUPS AFTER STEP   2
EACH F STATISTIC HAS   2 AND      146.0 DEGREES OF FREEDOM.

                GROUP        1            2

   GROUP

       2                804.16
                        0.0000

       3                1458.8          112.20
                        0.0000          0.0000
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                                  Page   8
19:20:09  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *


AT STEP   3, PW        WAS INCLUDED IN THE ANALYSIS.

                                    DEGREES OF FREEDOM  SIGNIF.   BETWEEN GROUPS
WILKS' LAMBDA          0.02492       3     2      147.0
EQUIVALENT F           257.854             6      290.0  0.0000


---------------- VARIABLES IN THE ANALYSIS AFTER STEP   3 ----------------
```

VARIABLE	TOLERANCE	F TO REMOVE	WILKS' LAMBDA
SW	0.7475999	55.036	0.04383
PL	0.7351089	38.979	0.03832
PW	0.6700620	35.822	0.03723

---------- VARIABLES NOT IN THE ANALYSIS AFTER STEP 3 ----------

VARIABLE	TOLERANCE	MINIMUM TOLERANCE	F TO ENTER	WILKS' LAMBDA
SL	0.3965791	0.3964933	4.7397	0.02338

F STATISTICS AND SIGNIFICANCES BETWEEN PAIRS OF GROUPS AFTER STEP 3
EACH F STATISTIC HAS 3 AND 145.0 DEGREES OF FREEDOM.

GROUP	1	2
2	692.77 0.0000	
3	1376.1 0.0000	131.92 0.0000

* *

AT STEP 4, SL WAS INCLUDED IN THE ANALYSIS.

		DEGREES OF FREEDOM		SIGNIF.	BETWEEN GROUPS
WILKS' LAMBDA	0.02338	4	2	147.0	
EQUIVALENT F	199.443		8	288.0	0.0000

---------- VARIABLES IN THE ANALYSIS AFTER STEP 4 ----------

VARIABLE	TOLERANCE	F TO REMOVE	WILKS' LAMBDA
SL	0.3965791	4.7397	0.02492
SW	0.6435380	24.906	0.03147
PL	0.3964933	38.032	0.03573
PW	0.6551095	27.298	0.03224

F STATISTICS AND SIGNIFICANCES BETWEEN PAIRS OF GROUPS AFTER STEP 4
EACH F STATISTIC HAS 4 AND 144.0 DEGREES OF FREEDOM.

GROUP	1	2
2	552.04 0.0000	

3	1092.8	103.23
	0.0000	0.0000

F LEVEL OR TOLERANCE OR VIN INSUFFICIENT FOR FURTHER COMPUTATION.

SUMMARY TABLE

STEP	ACTION ENTERED	REMOVED	VARS IN	WILKS' LAMBDA	SIG.	LABEL
1	PL		1	.05940	.0000	petal length
2	SW		2	.03723	.0000	sepal width
3	PW		3	.02492	.0000	petal width
4	SL		4	.02338	.0000	sepal length

CLASSIFICATION FUNCTION COEFFICIENTS
(FISHER'S LINEAR DISCRIMINANT FUNCTIONS)

SPECIES =	1	2	3
SL	19.61905	12.46482	9.807505
SW	26.19068	8.997175	5.231723
PL	-13.70145	7.385376	14.31615
PW	-18.60743	5.568353	20.56174
(CONSTANT)	-82.79144	-69.89760	-101.6665

CANONICAL DISCRIMINANT FUNCTIONS

FCN	EIGENVALUE	PCT OF VARIANCE	CUM PCT	CANONICAL CORR	AFTER FCN	WILKS' LAMBDA	CHISQUARE	DF	SIG
					: 0	0.0234	546.484	8	0.0000
1*	32.0777	99.09	99.09	0.9848	: 1	0.7733	37.399	3	0.0000
2*	0.2931	0.91	100.00	0.4761	:				

* MARKS THE 2 CANONICAL DISCRIMINANT FUNCTIONS REMAINING IN THE ANALYSIS.

STANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
SL	-0.39790	0.09856
SW	-0.54879	0.68806
PL	0.92044	-0.47046
PW	0.58602	0.61812

STRUCTURE MATRIX:

POOLED WITHIN-GROUPS CORRELATIONS BETWEEN DISCRIMINATING VARIABLES
AND CANONICAL DISCRIMINANT FUNCTIONS

(VARIABLES ORDERED BY SIZE OF CORRELATION WITHIN FUNCTION)

	FUNC 1	FUNC 2
PL	0.70240*	0.16393
SW	-0.11948	0.84827*
PW	0.63408	0.74861*
SL	0.21028	0.31179*

UNSTANDARDIZED CANONICAL DISCRIMINANT FUNCTION COEFFICIENTS

	FUNC 1	FUNC 2
SL	-0.7387907	0.1829974
SW	-1.615583	2.025554
PL	2.122543	-1.084885
PW	2.863515	3.020374
(CONSTANT)	-2.172296	-6.800889

CANONICAL DISCRIMINANT FUNCTIONS EVALUATED AT GROUP MEANS (GROUP CENTROIDS)

GROUP	FUNC 1	FUNC 2
1	-7.60132	0.21571
2	1.84638	-0.73710
3	5.75494	0.52140

TEST OF EQUALITY OF GROUP COVARIANCE MATRICES USING BOX'S M

THE RANKS AND NATURAL LOGARITHMS OF DETERMINANTS PRINTED ARE THOSE OF THE GROUP COVARIANCE MATRICES.

GROUP LABEL	RANK	LOG DETERMINANT
1	4	-13.067360
2	4	-10.546312
3	4	-8.677356
POOLED WITHIN-GROUPS COVARIANCE MATRIX	4	-9.723919

BOX'S M	APPROXIMATE F	DEGREES OF FREEDOM		SIGNIFICANCE
152.84	7.3422	20,	77566.8	0.0000

CLASSIFICATION RESULTS -

ACTUAL GROUP	NO. OF CASES	PREDICTED GROUP MEMBERSHIP 1	2	3
GROUP 1	50	50	0	0
		100.0%	0.0%	0.0%
GROUP 2	50	0	48	2
		0.0%	96.0%	4.0%

```
GROUP       3              50           0           1          49
                                       0.0%        2.0%       98.0%

PERCENT OF "GROUPED" CASES CORRECTLY CLASSIFIED:  98.00%


CLASSIFICATION PROCESSING SUMMARY

       150 CASES WERE PROCESSED.
         0 CASES WERE EXCLUDED FOR MISSING OR OUT-OF-RANGE GROUP CODES.
         0 CASES HAD AT LEAST ONE MISSING DISCRIMINATING VARIABLE.
       150 CASES WERE USED FOR PRINTED OUTPUT.
 11-Mar-94 SPSS RELEASE 4.1 FOR VAX/VMS                                  Page  13
19:20:10  WESTERN WASHINGTON UNIVERSITY  on NESSIE            VMS V5.4


Preceding task required .90 seconds CPU time;  2.43 seconds elapsed.

  20  execute

Preceding task required .01 seconds CPU time;  .01 seconds elapsed.

  21  finish

    21 command lines read.
     0 errors detected.
     0 warnings issued.
     1 seconds CPU time.
     3 seconds elapsed time.
       End of job.
```

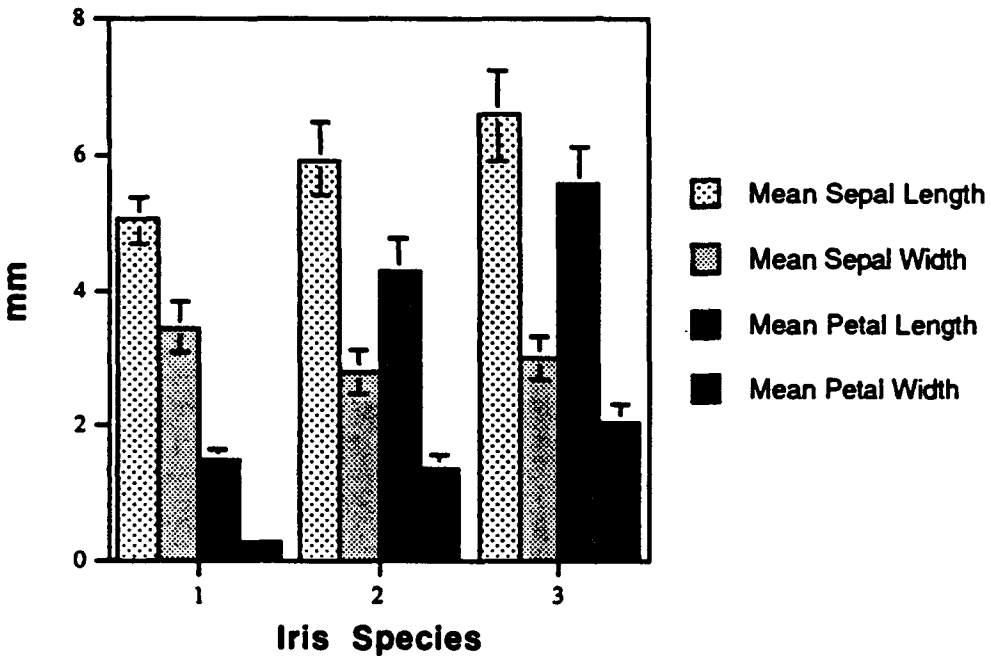
Appendix B. Summary of Iris Data

Take-Home Final

John Lenth
ENVR 451
March 13, 1994

IRIS DATA ANALYSIS

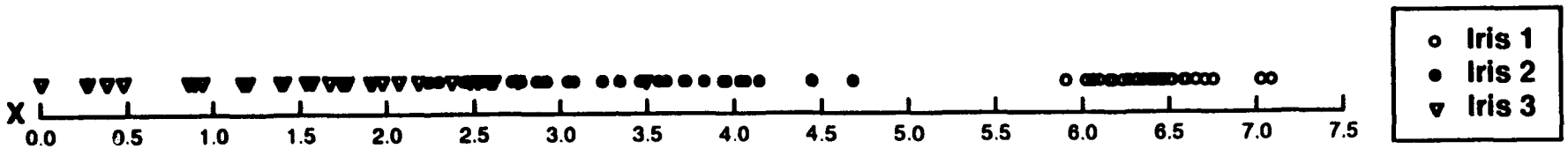
Bar Graph: Mean sepal length, sepal width, petal length and petal width with error bars representing the standard deviation.



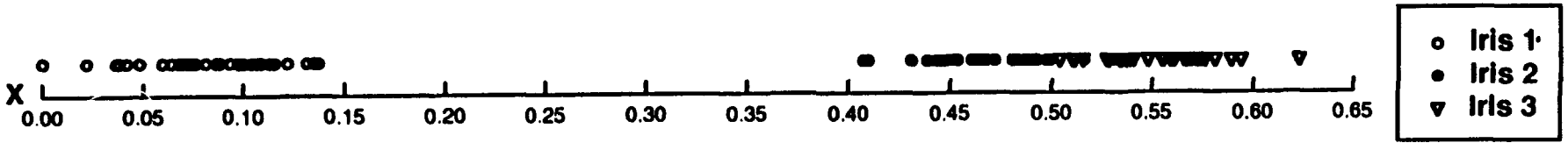
Discussion:

The mean for the following parameters increases from iris species 1 to 3; sepal length, petal length, and petal width. The mean sepal width appears to be approximately the same for the three iris species. In general, iris species 1 appears to have the smallest petals and sepals while iris species 3 has the largest. However, the error bars depicting the standard deviations for each parameter suggest that all of the observed differences may not be significant. For example, the error bars show considerable overlap exists between species 2 and 3 for mean sepal length. The above graph suggests that all the measurements taken on the iris species except sepal width are positively correlated (i.e. all the parameters tend to increase together).

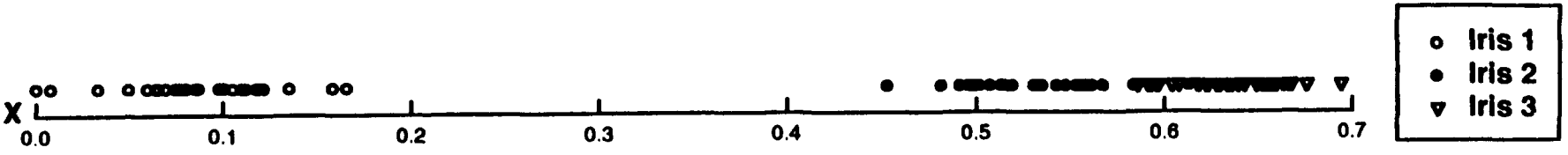
Single Axis Polar Ordination based on Euclidean Distance (99 SUs total, 33 SUs/Iris species)



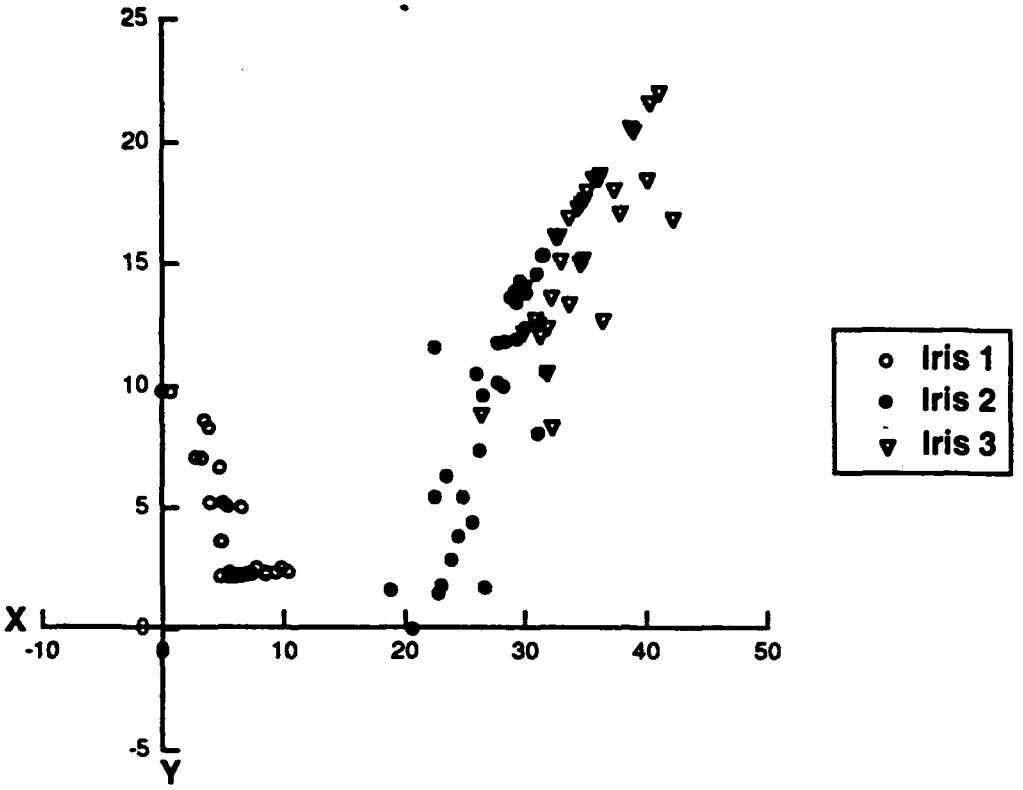
Single Axis Polar Ordination based on Relative Euclidean Distance (99 SUs total, 33 SUs/Iris species))



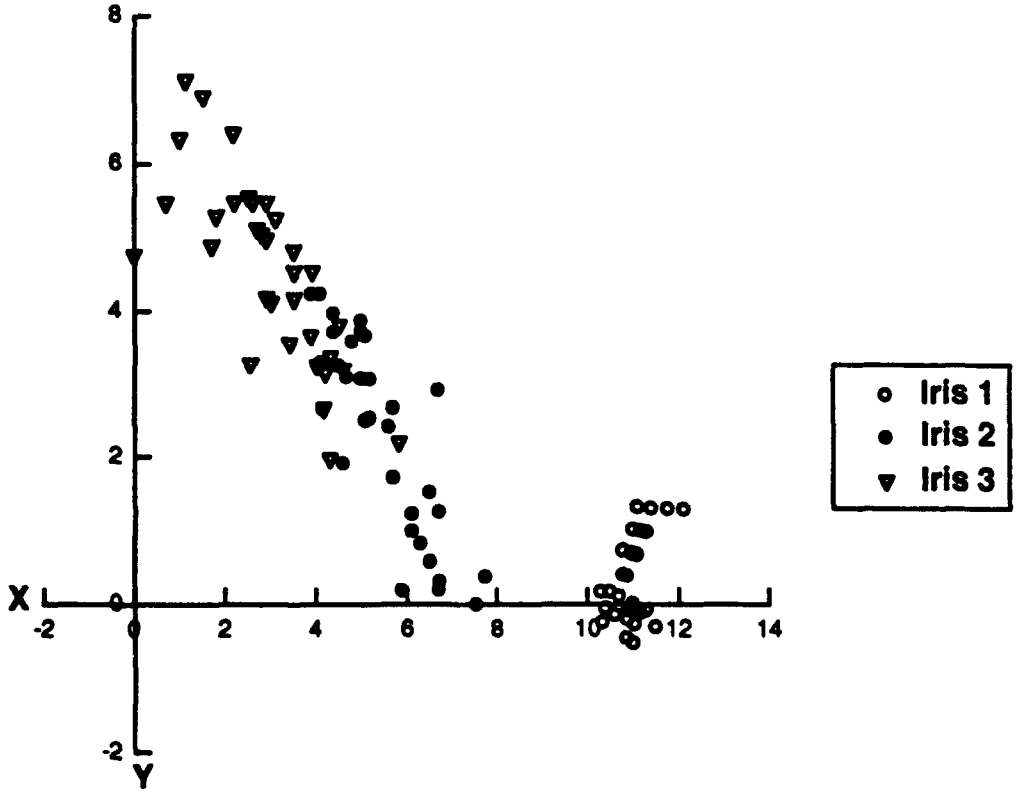
Single Axis Polar Ordination based on Mean Absolute Distance (99 SUs total, 33 SUs/Iris species)



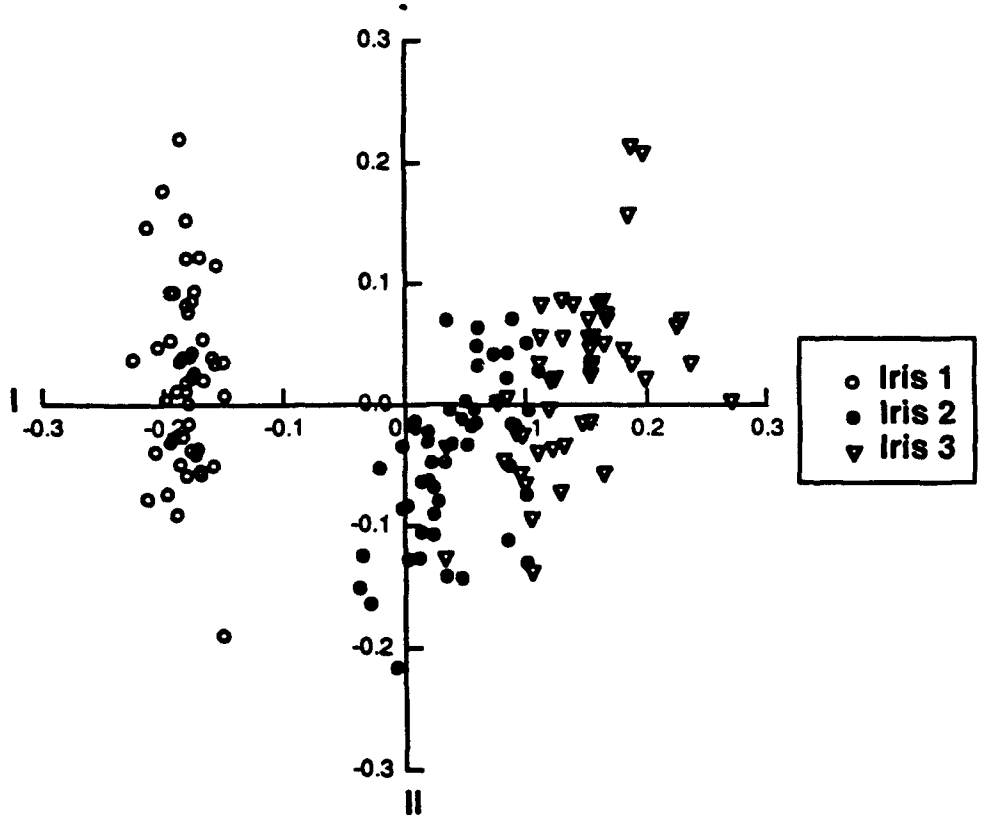
Polar Ordination based on Percent Dissimilarity (99 SUs total, 33 SUs/Iris species)



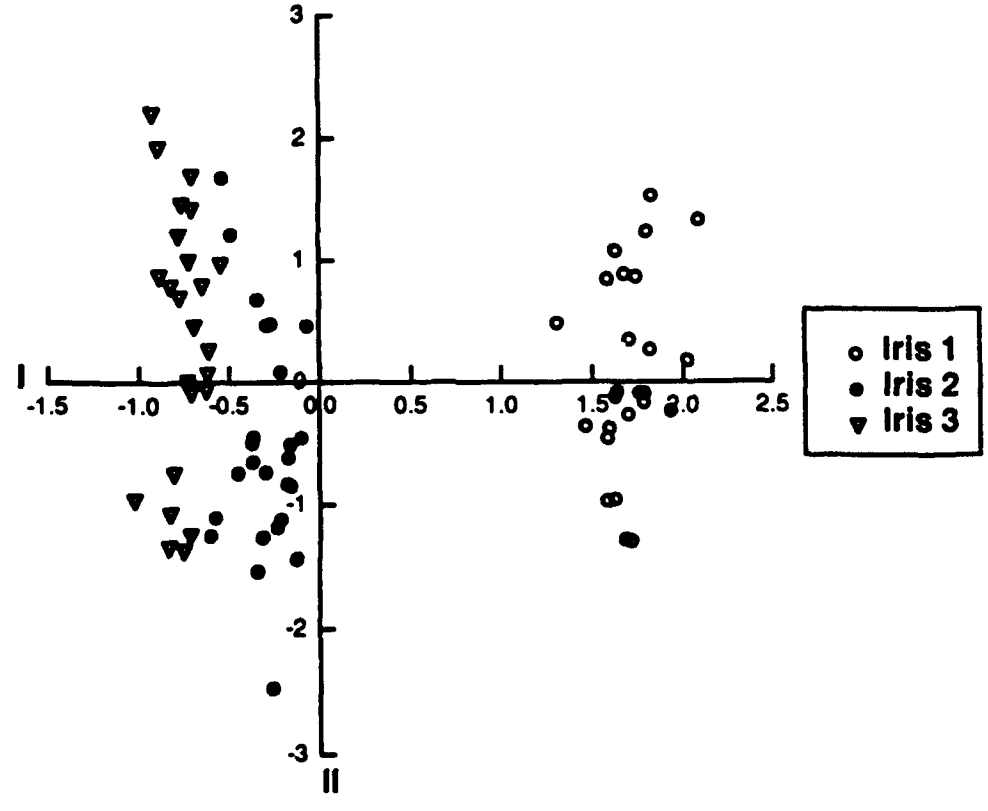
Polar Ordination based on Absolute Distance (99 SUs total, 33 SUs/Iris species)



PCA - Components I and II (150 SUs total, 50 SUs/Iris species)



COA - Components I and II (75 SUs total, 25 SUs/Iris species)



Discussion: Ordination Results

All the polar ordination techniques effectively demonstrated that differences do exist between the three iris species based on at least one of the parameters measured in the iris data set. The data points from iris species 1 were generally well separated along the x-axis from the data points corresponding to iris 2 and iris 3. While the data points from iris 2 and 3 did overlap in some areas, they still appeared to be enough distance between the majority of points to conclude that differences existed between these two species as well. Separation of iris 3 from iris 1 and 2 was observed along both the x-axis and the y-axis in polar ordinations with multiple axis. The separation of iris species along the x-axis generally reflected a gradient in petal lengths while the y-axis was more weakly associated with differences in sepal widths or lengths.

PCA and polar ordination appeared to reflect the same major trends in the data. For example, PCA also showed a pronounced separation of iris 1 from iris 2 and 3 on component I and a less pronounced separation of iris 2 and 3. This separation of data on component I was similarly associated with a gradient of petal lengths and accounted for 72.4% of the total variation in the data. Component II was less useful in identifying differences between the species because the data points for all three iris species overlapped significantly. This most likely occurred because component II is associated with sepal width which appears to be approximately the same for all iris species even though it has high overall variability (i.e. component II accounts for 22.9% of the total variation).

The results obtained using COA were nearly identical to those obtained using PCA. The same strong separation of the data points from iris 1 from those of iris 2 and 3 was observed on component I even though their position relative to the origin was reserved. Likewise, data points from iris 2 and 3 were more weakly separated and component II was of little use in identifying differences in the iris species.

Confirmatory Satistics: One way ANOVA and Tukey test for sepal length by iris species

Note : The following transformation was used to correct for heterogeneity of variances:
Transformed Data = Log_{10} (sepal length + 1)

H_O: The mean sepal length does not differ between iris species
H_A: The mean sepal length does differ between iris species
$\alpha = 0.05$

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	2	.2517	.1259	112.1700	.0000
WITHIN GROUPS	147	.1650	.0011		
TOTAL	149	.4167			

Conclusion: Reject H_O

TUKEY TEST: All pairs of iris species were significantly different at the 0.05 level.

Confirmatory Satistics: One way ANOVA and Tukey test for sepal width by iris species

H_O: The mean sepal width does not differ between iris species
H_A: The mean sepal width does differ between iris species
$\alpha = 0.05$

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	2	11.3449	5.6725	49.1600	.0000
WITHIN GROUPS	147	16.9620	.1154		
TOTAL	149	28.3069			

Conclusion: Reject H_O

TUKEY TEST: All pairs of iris species were significantly different at the 0.05 level.

Confirmatory Satistics: One way ANOVA and Tukey test for petal length by iris species

Note : The following transformation was used to correct for heterogeneity of variances:
Transformed Data = Log_{10} (petal length + 1)

H_O: The mean petal length does not differ between iris species
H_A: The mean petal length does differ between iris species
$\alpha = 0.05$

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	2	4.9694	2.4847	1902.9132	.0000
WITHIN GROUPS	147	.1919	.0013		
TOTAL	149	5.1614			

Conclusion: Reject H_O

TUKEY TEST: All pairs of iris species were significantly different at the 0.05 level.

Confirmatory Satistics: One way ANOVA and Tukey test for petal width by iris species

Note : The following transformation was used to correct for heterogeneity of variances:
Transformed Data = Log_{10} (petal width + 1)

H_O: The mean petal width does not differ between iris species
H_A: The mean petal width does differ between iris species
$\alpha = 0.05$

ANALYSIS OF VARIANCE

SOURCE	D.F.	SUM OF SQUARES	MEAN SQUARES	F RATIO	F PROB.
BETWEEN GROUPS	2	3.9110	1.9555	1388.2919	.0000
WITHIN GROUPS	147	.2071	.0014		
TOTAL	149	4.1181			

Conclusion: Reject H_O

TUKEY TEST: All pairs of iris species were significantly different at the 0.05 level.

Discussion: Summary of Results

The ANOVA results all showed that at least one of the iris species is significantly different for all the parameters measured. Furthermore, the Tukey tests showed that no two pairs of species were statistically the same for any of the measurements. Thus, the three iris species appear to differ significantly in relation to all the parameters measured.

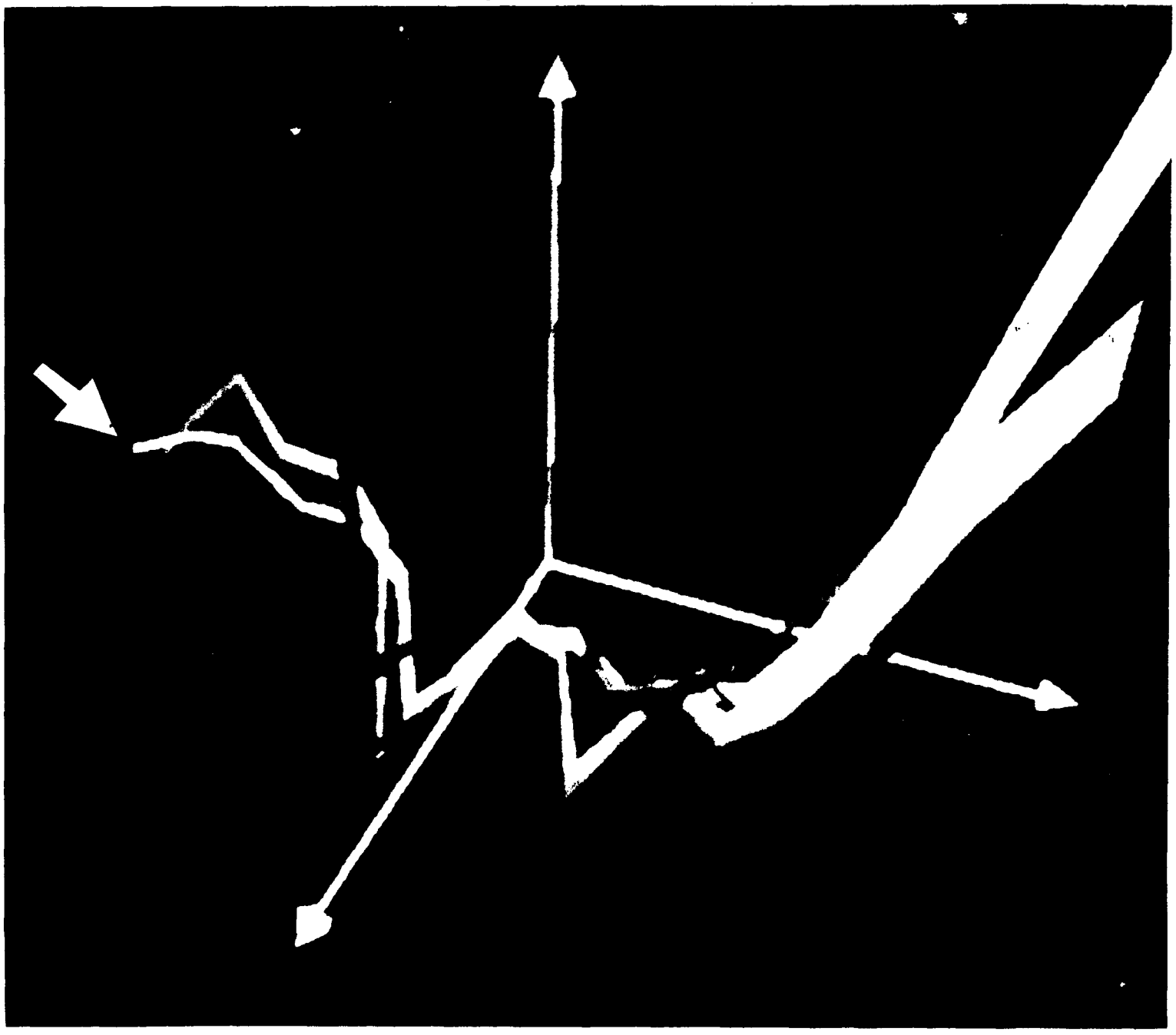
While it is important to know that statistically "significant" differences exist between the species, the ordination techniques were more useful for identifying the major patterns present in the data. For example, because petal length was consistently associated with the axis describing the most variability in the data, the ordination techniques all clearly showed that petal length was the most important parameter in differentiating the iris species . The relative magnitude of the difference between species could then be assessed by examining the distance between clusters of data points for the different iris species. If there was significant scatter in the data points for one iris species along either axis relative to the other species, it can be assumed the species may have a wider response range for one or more of the variables measured.

The ordination techniques were also able to show when similarities existed in the data. The large amount of overlap observed in the data points on component II of both PCA and COA suggests that sepal width was not as important in defining differences between the iris species despite the results of the ANOVA and Tukey tests. However, this illustrates a problem one may encounter when using ordination. A variable may have so much random variability that more meaningful patterns are lost in all the resultant "noise" that is generated in the data set. The bar graph on page one does suggest that other interesting patterns may have been present in the data that were left out or made less visible due to the influence of sepal width on the axes generated by the different ordination techniques. Other potential patterns might be more clearly elucidated by examining component III of PCA or by trying the ordinations without sepal width.

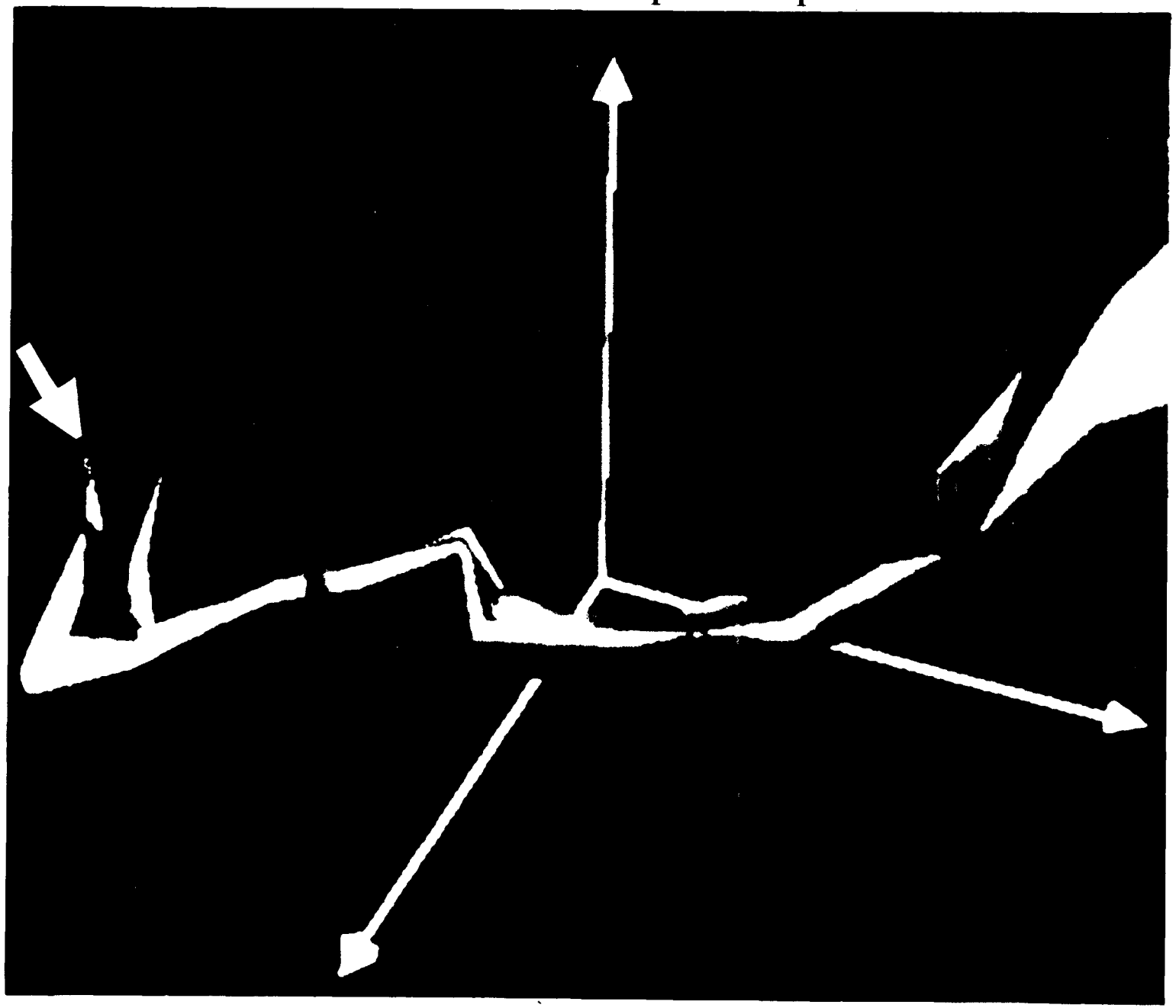
APPENDIX C

Color Reproductions of Space-time Worms

Jet-A PCA:C Projection Space-Time Worm



Jet-A Ankistrodesmus:Small Daphnia Space-Time Worm



APPENDIX D

M.S. in Environmental Science
Theses, 1991-1994

Draft Copy

Master's Thesis

Comparison of the Biodegradation of Water Soluble Components in Jet Fuel Using the Standardized Aquatic Microcosm (SAM) and the Mixed Flask Culture Microcosm (MFC)

April J. Markiewicz

Institute of Environmental Toxicology and Chemistry

Western Washington University

ES Bldg., Room 518. ext. 6137.

Draft Copy

Master's Thesis

Comparison of the Biodegradation of Water Soluble Components in Jet Fuel Using the Standardized Aquatic Microcosm (SAM) and the Mixed Flask Culture Microcosm (MFC)

April J. Markiewicz

Institute of Environmental Toxicology and Chemistry

Western Washington University

ES Bldg., Room 518. ext. 6137.

Abstract

The Standardized Aquatic Microcosm (SAM), a synthetic assemblage of organisms derived from laboratory cultures, was used in comparison to the Mixed Flask Culture Microcosm (MFC), derived from natural sources. Degradation rates and biodegradation products of water soluble components in jet fuel were monitored to evaluate whether the functional dynamics were similar in the two microcosms; independent of species diversity and trophic level complexity.

The SAM microcosms were used for the analysis of 1%, 5%, and 15% water soluble fraction (WSF) treatments of JP-8, and the MFC microcosms were used for the 1%, 5%, and 15% WSF treatments of Jet-A. Additional 15% WSF treatments were conducted on the 0% and 15% SAM and MFC treated microcosms to determine whether degradation rates would be increased due to the selective adaptation of hydrocarbon utilizing microbial populations.

Component degradation products and metabolites were monitored using Purge & Trap/Gas Chromatography. In both microcosms the concentration of the hydrocarbon class of compounds in the water soluble fraction determined the degradation rates for that class of compounds, rather than individual hydrocarbon component concentrations; initial structural and functional conditions in the microcosms determined degradation rates and persistence; both microcosm systems display the same patterns in degradation and metabolite production dynamics; only the SAM displayed increased rates of hydrocarbon degradation in the re-treated microcosms; and that metabolites from refractory hydrocarbon degradation appeared throughout the experiments.

Key Words: Microcosms, jet fuel, degradation rates.

This research is supported by USAFOSR Grant No. AFOSR-91-0291 DEF.

Introduction

Microbial degradation is the primary mechanism responsible for mediating the toxicity, persistence, bioavailability, and bioaccumulation of petroleum hydrocarbons in the environment (Atlas, 1981; Atlas and Bartha, 1993; Zobell, 1946, 1950). However, the importance and role of these microbial degradative processes have not been included in previous assessments of petroleum impacts to aquatic ecosystems (Gibson, 1977; Saunders, 1977). The development of multispecies ecotoxicological testing procedures, or microcosms, is a progressive step towards providing simpler, replicable, and readily standardized testing systems that can be used to facilitate the study of the functional processes mediated by microorganisms in the context of ecosystem-level dynamics. The objectives of this study were to evaluate microbial community hydrocarbon degradation mechanisms, rates, and transformation products in two aquatic microcosms, the Mixed Flask Culture (MFC) microcosm (Leffler, 1980; Shannon and Anderson, 1989) and the Standardized Aquatic Microcosm (SAM) (ASTM E1366-91, 1991; Taub and Read, 1982). The purpose was to determine whether the two microbial communities display similar patterns in degradative rate responses and products, independent of petroleum fuel composition, microcosm type, microcosm species diversity, and trophic level complexity.

The global impacts of petroleum production and industry on atmospheric, aquatic, and terrestrial environments have increased several orders of magnitude during the last fifty years. The frequency and quantity of petroleum currently being transported throughout the world has resulted in greater discharges, spills, and emissions into the global ecosystem that has surpassed the ability of indigenous microbial communities to degrade and eliminate. In addition, the use and steady depletion of readily accessible sources of petroleum has caused increased efforts in the exploration for petroleum reserves in more remote wilderness areas and off-shore locations. The potential for these activities to contaminate and destroy ecologically fragile environments, is rapidly being realized. Atmospheric transport and deposition, coupled with these anthropogenic activities, have resulted in a global ecosystem where true hydrocarbon-free sites no longer exist (Vandermeulen, 1978).

The operational, accidental, or intentional release of hydrocarbon mixtures are at any concentration, deleterious to the plant and animal life associated with that environment. Most of these hydrocarbon contaminants are similar in structure to naturally occurring compounds and as a result are more hazardous, due to their ability

to biochemically interfere with the enzymatic mechanisms of induction and modes of action in the organisms (Gibson, 1977; O'Neill and Waide, 1981).

The physical effects of these compounds can be equally destructive and toxic. Higher molecular weight components will immediately coat, smother, and asphyxiate both organisms and plants by the physical blocking of airways and stomata. The indirect effects of these coatings will be manifested in the consumption of the hydrocarbons by organisms either by preening to remove the coating from their feathers, hair or skin, or by the ingestion of plants or algae contaminated with the hydrocarbons (Beer, 1968; Bourne, 1968).

Lower molecular weight components will penetrate directly through pores or stomata and indirectly through uptake by root systems, diffusion, and dissolution of fatty molecules in the outer membranes of the organisms and plants (Browning, 1953). These lighter petroleum hydrocarbon components dissolve in the plasma membrane in both plants and animals and create spaces within the membrane structure by displacing the fatty molecules. The damage to the plasma membrane increases permeability and may cause cell contents to leak into the intercellular spaces, or for materials to move from within the cells into the hydrocarbon fraction (Vandermeulen and Ahern, 1976). Localized regions of chlorosis and necrosis may be prevalent on the leaf structure or the entire plant may be destroyed. Processes involving transpiration and photosynthesis may also be impaired and reduced, while respiration may be increased or decreased depending on the plant species involved (Baker, 1971; Vandermeulen and Ahern, 1976).

The refractory compounds composed of the high molecular weight polycyclic aromatics and asphaltenes will remain in the environment at the site of the release for periods of time ranging from months to several years. The potential for these compounds to re-contaminate and affect the biota will be dependent on the resuspension and trophic cycling events occurring within that environment.

The release of petroleum hydrocarbons to the environment, whether in crude or refined form will cause effects that are physically damaging and toxicologically destructive over a time span of several years. The potentiation of these effects will be dependent on a complexity of direct and indirect, biotic and abiotic processes and dynamics. Specifically, these will involve the indigenous microbial communities and the physical, structural, and functional components and processes of the ecosystem, interacting with the physicochemical properties of the petroleum hydrocarbons (Gibbons, 1977; Giddings, 1983). In prior studies, the focus has been primarily on observations and measurements taken at the actual spill or discharge site and on laboratory derived dose-response relationships where selected "representative" or

"sensitive" laboratory organisms were exposed to serial dilutions of petroleum mixtures under controlled environmental conditions (Connell and Miller, 1984). The use of microcosms is means of integrating these methods into simpler, cost effective, replicable testing systems that will enable the measurement of the potential ecosystem-level fate and effects of toxicants under more controlled, statistically robust conditions (Giddings, 1983; Hammons et al., 1981; Kroer and Coffin, 1992; Leffler, 1980; Shannon et al., 1986; Suter II, 1993; Taub, 1984; Wilkes, 1977).

Microcosms

The ability to assess the effects of petroleum hydrocarbons on ecosystem functional processes and biological structural components, prior to their release to the environment, will allow valid predictions of their potential for ecological risk to be formulated. In turn, regulatory decision-making will be improved for the greater protection of dwindling environmental resources (Bartell et al., 1992; Suter II, 1993). The use of microcosms is an attempt to integrate these ecosystem-level functional and structural processes with chemical hazard assessments to facilitate predictions of ecological risk (Giddings, 1981; Hammons et al., 1981; Shannon and Anderson, 1989; Suter II, 1993; Taub, 1983). The design of the microcosms to be smaller, trophically simpler, replicable, and standardized testing systems allows closer examination of specific relationships and interactions in determining direct responses to direct effects. In addition, it reduces the natural variances associated with analyzing any complex system that could potentially diffuse or hide effect responses. It also allows the comparison of results obtained in different laboratories, without microcosm design differences becoming a factor (Suter II, 1993). As with most experimental designs and testing systems, the use of microcosms in environmental risk assessments is a compromise to obtain the most information of ecosystem processes, in relation to chemical fates and toxicological effects, without having to conduct field tests and still maintain some acceptable level of realism.

Microcosms were originally developed and used for studies of population dynamics, species interactions, and community structural and functional relationships (Beyers, 1963; Leffler, 1980; Taub, 1984). In these early tests, many of the processes and population interactions that occur in natural ecosystems were demonstrated to occur in the synthetically created microcosms including photosynthetic production and respiration dynamics, algal competition and succession, grazing effects, and nutrient cycling (Beyers, 1963; Giddings, 1983; Shannon and Anderson, 1989; Taub, 1980, 1983). These preliminary results encouraged the rapid development,

modification and use of microcosms in chemical fate and toxicological effects tests to provide greater dimensionality and realism beyond the level of the single species toxicity tests being conducted (Giddings, 1981).

Current microcosm toxicological tests conducted are similar to their predecessors, but with less systematic reviews, evaluations, and comparisons made of the different methodologies. The novelty of conducting these experiments has resulted in experimental designs that in many instances, are inappropriate for the hypothesis being tested. The hypothesis is either not explicitly stated, or is severely limited in ecological significance. The analytical parameters focus primarily on the biological structural components with a few physical parameters included, such as pH, dissolved oxygen, conductivity, and alkalinity. Species are identified and enumerated during the course of the experiment to determine changes in diversity and abundance patterns, with survival used as the endpoint to indicate organism response to the toxicant effect. The premise of using this approach is that because ecosystems are so complex that by focusing on the functions, interactions, and responses of the individual parts would, when combined together, reveal and explain whole ecosystem dynamics (O'Neill and Waide, 1981). Ecosystems are not, however, the sum of their individual components, nor can measurements of a few parameters be extrapolated to infer natural ecosystem responses and properties (O'Neill and Waide, 1981).

The limitations of using components to assess effects to the whole ecosystem are apparent. Community structure is frequently altered in the environment by natural or anthropogenic stress events but is not necessarily reflected in alterations in ecosystem functional processes. Similarly, alterations in ecosystem functions may not necessarily produce changes in community structure (Matthews et al., 1982). O'Neill and Giddings (1979) showed that the elimination of an algal population had little effect on photosynthetic productivity, due to the release of other algal populations from competitive inhibition and their subsequent utilization of the freed nutrients. Conversely, the elimination of certain microbial communities that are responsible for the degradation and decomposition of organic matter in a system will result in profound changes in nutrient cycling. These changes will not be displayed by immediate alterations in the higher trophic level populations that are usually monitored (Sheehan, 1984a).

The objectives of progressive ecosystem-level tests must become more explicit. The parameters measured must be extended to include more complex functional components beyond the standard physical data of pH, dissolved oxygen, conductivity, and nutrient levels. Community metabolism rates that include photosynthesis and

respiration ratios, total CO_2 efflux, biochemical rates, ATP concentrations, chlorophyll concentrations, nutrient cycling, dissolved oxygen concentrations, pH, substrate decomposition rates, toxicant degradation rates, bioaccumulation rates, and accumulation rates of metabolic by-products become especially important parameters to measure when comparing between microcosm systems or to field tests (O'Neill and Waide, 1981; Sugiura, 1992). The utilization of these processes with the biological components may provide the necessary insights to determine whether there are specific patterns in the rate responses that are the direct (or indirect) result from exposures to specific chemical classes of compounds (Sheehan, 1989).

The advantages of using rate responses are that they display similar first order kinetics, logistic growth curves, or proportionality functions that are comparable both within and between systems (Alexander, 1985). Current testing strategies have not been extended beyond simple P/R ratios and nutrient cycling to determine whether the rate responses and intensities in one microcosm system are comparable to those determined in another microcosm or to the ecosystem they are meant to simulate (Sugiura, 1992). Instead of attempting to simulate ecosystem complexity and realism, microcosm designs should be selected that provide the effect responses important for a particular combination of chemical stressor and ecosystem type to be investigated (Suter II, 1993).

The two aquatic microcosms selected for this study were the Mixed Flask Culture (MFC) microcosm, developed by Leffler (1984) and later modified by Shannon and Anderson (1989), and the Standardized Aquatic Microcosm (SAM), developed by Taub and Read (1982). The major advantage of using these "generic" microcosms is that they are both standardized in terms of species composition and are constructed to be ecologically similar at the initiation of the experiment (Giddings, 1983; Suter II, 1993). The organisms used in the two microcosms have distinct interspecific and intraspecific interrelations and responses that are comparable to those present in the environment (Giddings, 1983; Leffler, 1980; Suter, 1993; Taub, 1984). However, both systems are synthetically assembled to produce taxonomically simple or generic communities and in SAM methodology, are gnotobiotic or completely defined populations.

The MFC and the SAM have the same artificial sediment, the same chemically defined media, and an assemblage of organisms representing different trophic levels inoculated into them. The primary differences between the MFC and the SAM microcosms were the source of their species assemblages, their level of "realism" in terms of species diversity and complexity, and their size. The MFC populations were collected locally from natural aquatic environments, combined together in a chemically defined

medium, and allowed to "co-adapt" for a predetermined length of time. The resulting artificial community was used as the stock community inoculum in the construction of the 1.0 L microcosms. The MFC microcosms have the same major taxa as the SAM's but are generally more species rich, with greater species diversity and trophic level complexity. In the SAM the populations are axenic, laboratory cultured organisms that are individually inoculated at known densities directly into the 3.8 L microcosm containing the chemically defined medium. Its simpler composition precludes species diversity and trophic complexity.

An advantage of the MFC microcosm protocol is the "realism" inherent to the more natural assemblage that potentially enables greater extrapolation of test results to the ecosystem from where it was derived (Leffler, 1980; Shannon and Anderson, 1989). An advantage of the SAM protocol is the reproducibility and statistical robustness of the test results that presumably enables their extrapolation to many types of ecosystems. Another advantage is the presumed sensitivity of "new" versus "mature" SAM microcosms. The new microcosms are thought to display structural alterations or dose-response patterns with greater sensitivity and amplitude immediately following toxicant exposure than mature, aged microcosms (Kindig et al., 1983; Taub, 1984).

The underlying assumption of using generic microcosms for ecosystem-level toxicity testing is that all ecosystems display the same patterns and behaviors in their structural and functional relationships and processes. That there exists universal ecosystem properties and universal patterns of responses to stress (Giddings, 1983). Ecological realism and complexity are sacrificed for the purpose of discerning and defining these generic ecosystem-level structural and functional processes that could be applied to all ecosystems.

The adaptation of these microcosms for analyzing microbial community interactions and functional responses after exposure to jet fuel contaminants required several considerations and modifications. One modification was the selection of an appropriate parameter to measure that would indicate the rate responses of the hydrocarbon transformation and mineralization processes. The sensitivity of the parameter measured to resolve changes in the rate responses and intensities were also considered as well as the analytical accuracy by which the parameters could be determined. Another modification was the time frame within which the measurements were taken (Kroer and Coffin, 1992; Saunders, 1977; Sugiura, 1992). Of these factors, the sampling frequency was considered to be one of the most critical. The dynamics of microbial population growth and substrate utilization rates mandate a sampling frequency that lies within the longevity of the organism's life cycle. Samples

collected at a frequency ranging from days to weeks will not reveal microbial degradative pathways and mechanisms that occur within a period of a few hours (Saunders, 1977).

Experimental Design

The MFC and SAM microcosm systems were used to monitor the rates of microbial degradation and metabolic by-products formed during the biodegradation of the hydrocarbon components in the water soluble fractions of two jet fuels. The parameters that were selected for analysis were the concentrations of the individual hydrocarbon components as they were degraded and the concentrations and types of metabolic intermediates that were produced. All hydrocarbon analyses were conducted using a purge and trap concentrator system in conjunction with a gas chromatograph. Qualitative identifications and concentrations were determined for the hydrocarbon components in each of the jet fuels and the metabolic intermediates produced (USEPA, 1982).

The rates of biodegradation were calculated for each hydrocarbon component by regressing the decrease in concentrations through time and using the slopes as rates of degradation (Walker and Colwell, 1976a). The rates of hydrocarbon component degradation were compared within the microcosm experiments to determine the effects of concentration on degradation rates. The rates were also compared between microcosms to determine significant differences or similarities in microbial degradation rates and metabolic pathways. The intermediate compounds produced were compared as well to determine potential similarities in metabolic pathways.

The selection of the Mixed Flask Culture microcosm and the Standardized Aquatic Microcosm, that are different in species composition, structural complexity, size, and construction, was to evaluate and compare whether they display similar patterns of rate responses and intensities in the degradation of hydrocarbon components, when treated with the water soluble fraction of a jet fuel. Secondly, to determine whether the same rate patterns are repeated, when they are re-treated with a second water soluble fraction of jet fuel. Third, to determine whether the patterns are similar to responses and intensities observed in field studies. Finally, to determine whether microcosms must resemble real ecosystems, as closely as possible, to be valid models of ecosystem dynamics or whether ecosystem dynamics display similar patterns in rate responses independent of species composition and trophic complexity.

Jet Fuel

The jet fuels JET-A and JP-8 that were used in this study, are composed of complex mixtures of hydrocarbons that have been refined and blended to produce artificial formulations. The fuels are produced by blending certain percentages of naphtha (alkanes and aromatic hydrocarbons), gasoline (C_5 to C_9 alkanes, alkenes, cycloalkanes, and aromatics), and kerosene (n-dodecane, alkyl-benzene derivatives, naphthalene, and its derivatives) to meet commercial and military specifications (Riser-Roberts, 1992). The JET-A fuel is the most commonly used aviation fuel in the commercial sector. The JP-8 fuel is a new formulation recently created by the U.S. Air Force as a less toxic alternative to the JP-4 fuel currently being used by the military sector. The specific hydrocarbon concentrations and components within each jet fuel will vary with each manufactured lot. The source of the component mixtures is variable from lot to lot. The specific characteristics will be dependent on the geological and geographical origin of the initial crude oil as well as the nature of the cracking procedure used during the refining process (Connell and Miller, 1984; Perry, 1980; Riser-Roberts, 1992).

The jet fuel components belong to three major classes of hydrocarbons that are characterized by their chemical structure. The alkanes are the normal, straight chain paraffins or saturates and include the cyclic alkanes. The alkenes are the olefins or unsaturated nonaromatics and the aromatics are the mono-, naphthalene, and polycyclic aromatic compounds (Boesch et al., 1974). The specific chemical structures and the concentrations of these three classes of petroleum hydrocarbons in the jet fuels will determine their chemical properties. These properties will influence their solubility, volatility, toxicity, persistence, and resistance to rates of photochemical oxidation and microbial degradation in the environment (Davis, 1967; Riser-Roberts, 1992).

Hydrocarbon Chemistry and Toxicology

The alkanes are chains of carbon atoms with attached hydrogen atoms and may be simple straight chains (n-normal), branched (iso-, sec-, tert-), or have a simple ring configuration (cyclo-) (Figure 1). Low molecular weight alkanes have low boiling points and are highly volatile. They are slightly soluble in water and extremely soluble in fats and oils that enhances their rapid penetration through membranes and into tissues. Alkanes act primarily by solubilizing or emulsifying fats, mucous membranes, and cholesterols (Browning, 1953). In mammals alkanes have also been found to penetrate rapidly into the fatty cells of the myelin sheath that surround nerve fibers, dissolve the nerve cells and cause degeneration of the axon, interrupting the transference of nerve impulses (Manahan, 1989).

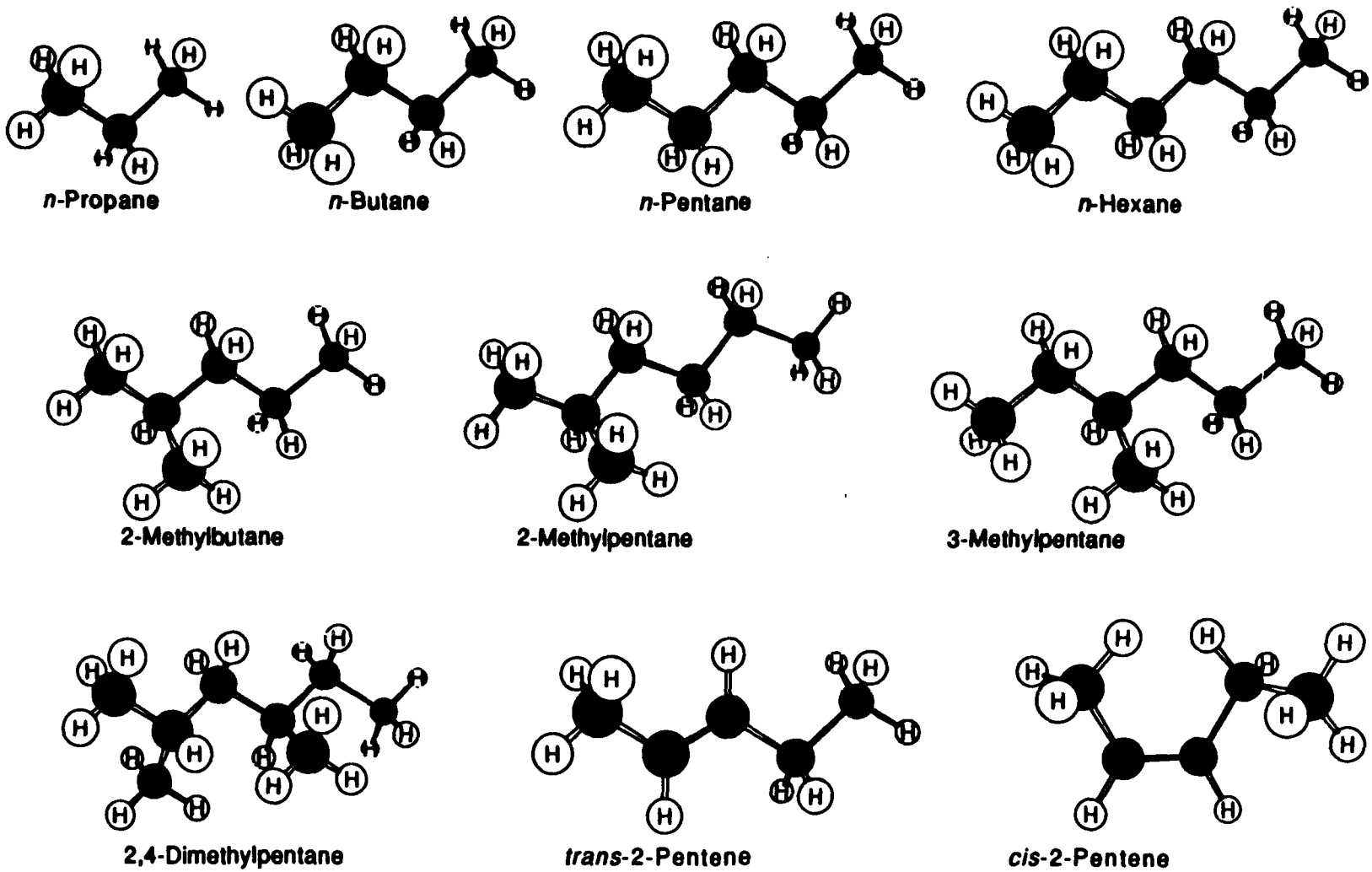


Figure 1. Chemical structures for *n*-alkanes, alkyl-substituted alkanes, and alkenes.

High molecular weight alkanes are exclusively lipophilic but are considered to be virtually non toxic though they may affect chemical communication and interfere with metabolic processes. Many of the same high molecular weight alkanes are produced biogenically and have been found occurring naturally in terrestrial plants, aquatic algae, and macrophytes as well as in all marine organisms (Boesch et al., 1974; Riser-Roberts, 1992). Some of the alkanes detected and identified in this study were *n*-propane, 2-methylbutane, *n*-pentane, 2-methylpentane, 3-methylpentane, *n*-hexane, 2,4-dimethylpentane, *n*-decane, *n*-dodecane, *n*-tridecane, and *n*-tetradecane (Figure 1).

The alkenes are also chains of carbon atoms with attached hydrogen atoms, but the chains contain carbon-carbon double bonds and are unsaturated in relation to the total possible number of attached hydrogen atoms, compared to an alkane of similar carbon chain length. The double bonds convey a planar configuration that allows the formation of the geometrical isomers *cis*- and *trans*- (Figure 1). Alkenes are generally more reactive due to the presence of the unsaturated double bond that provides a location for chemical attack not present in alkanes. They are present specifically in refined petroleum products, such as gasoline and aviation fuels. Alkenes undergo addition reactions which increase their chemical and metabolic capabilities in forming potentially more toxic metabolites. They can be transformed by three pathways. They can be polymerized to create long polyethylene chains, oxidized to form oxides that on hydrolysis can form glycols, and halogenated to form extremely toxic chlorinated and brominated hydrocarbon pesticides (Browning, 1953; Manahan, 1989).

In experimental animals the *cis*- isomer has been found to be an irritant and narcotic that causes damage to the liver and kidney. The *trans*- isomer has been found to cause weakness, tremors, and cramps due to its effects on the central nervous system as well as nausea and vomiting from adverse affects involving the gastrointestinal tract (Manahan, 1989). The alkenes detected in this study were tentatively identified as *cis*-2-pentene and *trans*-2-pentene and were believed to be the metabolic by-products from the degradation of the aromatic compounds.

The aromatics hydrocarbons have a basic six carbon atom ring configuration with six hydrogen atoms and three double bonds and are unsaturated in attached hydrogen atoms. The aromatic ring may occur in a single configuration as benzene, in two attached rings to form naphthalene, or in many attached rings as polycyclic aromatic hydrocarbons (PAH's) (Manahan, 1989). The aromatic ring structures may also have substituted methyl or more complex alkyl side chains as in the case of toluene, xylene, ethylbenzene, and propylbenzene (Figure 2). The substitution of the hydrogen atoms on

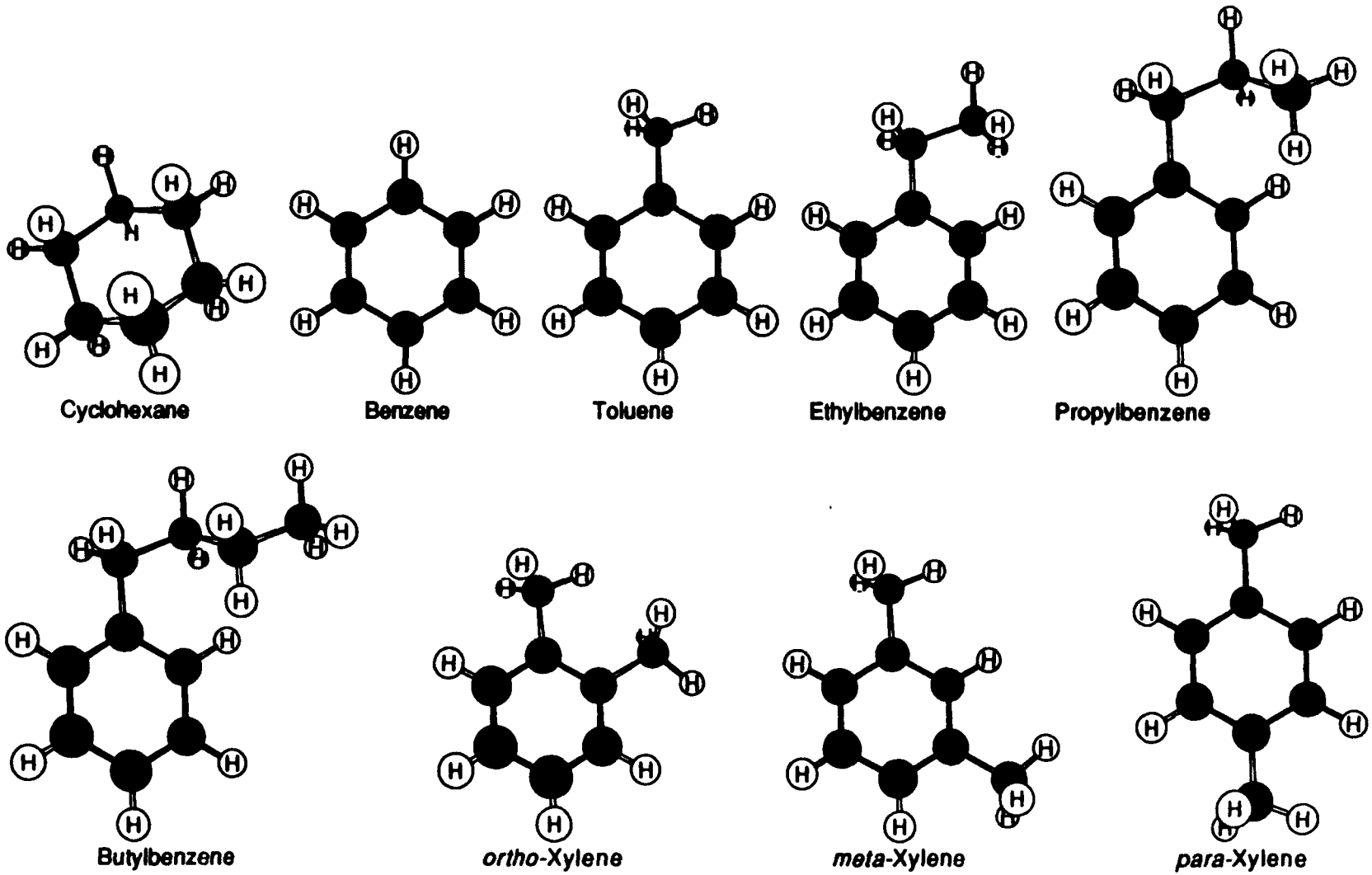


Figure 2. Chemical structures for cycloalkane, mono-aromatic, and alkyl-substituted aromatic compounds.

the aromatic ring, will alter the degree of polarity, lipophilicity, persistence, and toxicity of the compound (Rochkind, 1986; Manahan, 1991). The carbon-carbon, resonance stabilized bonds of the aromatic ring structure confers an increased stability to these compounds making them not only acutely toxic but some of the most persistent and carcinogenic in the environment (Brown, 1982; Manahan, 1989).

Benzene, toluene, and the three isomers of xylene are among the most common monocyclic aromatic chemicals found in jet fuel (Davis, 1967; Moore and Ramamoorthy, 1984; Riser-Roberts, 1992). They have low molecular weight, low water solubility, mid-range octanol-water partition coefficients, high volatility and flammability, and their toxicological modes of action are narcosis (Manahan, 1991; Rappoport, 1967). Their structure, stability, and ability to be both slightly hydrophilic and lipophilic enhances their accessibility to more niches, species, biochemical pathways, and sites of action that accounts for the subsequent assessment of these compounds as priority pollutants by the USEPA in 1977 (Rochkind, 1986; Manahan, 1991; Moore and Ramamoorthy, 1984). As a result of their ecotoxicological significance, the biodegradation rates and metabolic by-products of these compounds were emphasized in this study.

Benzene (bp 80.1°C) is a potent narcotic, that affects the central nervous system. At high concentrations inhalation of air containing approximately 64 g/m^3 of benzene can be fatal within a few minutes and one tenth of that level can cause acute poisoning within an hour (Manahan, 1989). Exposure causes skin irritation, fluid accumulation in the lungs (edema), excitation, depression, and may eventually lead to respiratory failure and death. At lower concentrations benzene can cause blood abnormalities, lower white cell count, and damage bone marrow (Browning, 1953; Manahan, 1989). These toxicological effects have been attributed specifically to the *trans*-benzene-1,2-oxide intermediate formed during the eukaryotic oxidation of benzene (Manahan, 1989). Prokaryotic oxidation specifically forms *cis*-benzene-1,2-oxide (Gibson, 1976) (Figure 4).

The oxidative process involves the incorporation of the oxygen atom directly into the ring structure to form an epoxide. This epoxide intermediate, which is not immediately degraded, resides in the cell structures actively reacting with cell nucleophiles. Cellular damage to the blood, lymph, and bone marrow cells results and ultimately affects liver and kidney function (Manahan, 1991). The epoxide structure is eventually converted to phenol by a slower, nonenzymatic rearrangement process and is eliminated from the body. Approximately 16% is respired as benzene within the first five hours of exposure, while the remainder is released more slowly, in the form of

phenol sulfate, with the rate of release being dependent on the amount of benzene stored in the lipophilic tissues (EPA, 1986; Manahan, 1989). Benzene is of most concern due to its known association with the development of leukemic cancer in humans.

Toluene (bp 110.6°C) is also a narcotic but is more potent than benzene. At low concentrations it produces skin irritations and at higher levels affects blood cells, the liver, the kidney, and the central nervous system to cause headaches, nausea, and impaired coordination (Browning, 1953; Manahan, 1989). Toluene is less water soluble and more lipophilic than benzene. It rapidly penetrates membranes and transported to the site of action at greater concentrations that increases its potential for toxic effects (Kauss and Hutchinson, 1975). However, the rapid enzymatic degradation of toluene mediates the site concentration and reduces its potential toxicological effects (Berry and Brammer, 1977; Donahue et al; 1977, Kauss and Hutchinson, 1975). The mechanism involved in moderating these effects is the rapid oxidation of the methyl side-chain that is enzymatically more accessible than the more stable ring structure. The benzyl alcohol and benzoate intermediates formed are conjugated to hippuric acid and are rapidly eliminated. This metabolic pathway account for approximately 70% of the dose with the remainder being respired from the lungs unchanged as toluene (Rochkind et al., 1986; Manahan, 1989).

The xylenes *ortho, meta, and para* (bps 144.4°C, 139.1°C, and 138.3°C, respectively) act as narcotics on the central nervous system but to a much lesser extent than benzene and toluene. At high concentrations they cause headaches, impaired coordination, edema, and nausea. At lower concentrations they cause skin irritations, anemia, blood cell damage, and reduce blood platelets (Browning, 1953; Manahan, 1989). The double methylation of the xylenes make them virtually insoluble in water. They are very lipophilic with high octanol-water partition coefficients and the potential for rapid transport to the site of action. The toxicity of the xylenes is mediated by the rapid oxidation of one of the methyl substituted groups and the cleavage of the aromatic ring. The position of the second methyl group on the benzene ring determines the number of enzymatic steps in the degradation process, the specific metabolic pathway, the rate of degradation, and the potential for bioaccumulation (Berry and Brammer, 1977; Donahue et al., 1977; Evans et al., 1991; Kauss and Hutchinson, 1975; Perry, 1979; Worsey and Williams, 1975). The elimination of xylenes is primarily through the excretion of metabolites in the form of methyl hippuric acid that represents 95% of the absorbed dose, 1-2% as xylenols, and 3-5% respired from the lungs as unchanged xylene (Rochkind et al., 1986).

The toxicological effects of ethylbenzene, propylbenzene, and butylbenzene are intermediate between benzene and toluene. Their narcotic activity is slower, of longer duration and primarily depressant. Neither acute or chronic poisoning has been recorded in humans (Browning, 1953).

Microbial Degradation

Microbial communities in all ecosystems are composed of the same types of microorganisms that perform the same functional processes on which energy flows and nutrient cycling are critically dependent (Atlas and Bartha, 1993; Fenchel, 1977; Ford, 1993; Zobell, 1946). The utilization of hydrocarbons as a substrate for energy and growth by a microbial community will depend on the type of microorganisms, the similarity of the hydrocarbon chemical structure to natural substrates utilized by the microorganisms, nutrient availability, the frequency of the hydrocarbon exposure, its concentration, and the commensal and cometabolic relationships existing between the microbial populations (Atlas, 1981; Gibson, 1977).

Indigenous aquatic microbial communities are composed of autotrophic and heterotrophic prokaryotes that include algae, bacteria, viruses, fungi, molds, and yeasts (Atlas and Bartha, 1993; Brock et al., 1994; Ford, 1993; Riser-Roberts, 1992). The microbial community structure in each habitat will be determined by the requirements in each micro-environment for specific microbial processes and the abiotic constraints on the types of viable and active microorganisms at any given time. In addition, constraints by anthropogenic perturbations will also affect community structure (Ford, 1993).

The relative success of a microbial population will depend on its ability to selectively adapt and utilize a nutrient or xenobiotic and on its physiological rates of nutrient uptake, inherent metabolic rates, and growth rates (Brock et al., 1994: Gibson, 1977). The critical environmental factors that directly affect microbial metabolic degradative rates and growth rates are temperature, light, nutrient concentrations of nitrogen and phosphorus, and oxygen availability (Atlas and Bartha, 1993; Brock et al., 1994; Ford, 1993). In addition, the effects of latitude, season, and watershed hydrogeochemical and biological processes will mediate nutrient quality and availability. Trophic/food-web interactions, predation, and competition as well as wind and wave activity, depth, and pressure will further limit optimal conditions for any one microbial population, at any given time (Atlas, 1981, 1988, 1991; Brock et al., 1994; Focht, 1988; Wolfe, 1987).

In aquatic environments most microorganisms are found preferentially associated with organic particulate and dissolved matter or attached and growing on the surfaces, or enclosed in polysaccharide or chitin biofilms (Atlas and Bartha, 1993; Brock et al., 1994; Bull and Slater, 1982). The surfaces utilized may be inorganic or organic matter and may include soil or sediment particles, living or dead algal cells, or other organisms. The bacterial cells attach by excreting adhesive polysaccharides and use the biofilms to trap nutrient or xenobiotic substrates for growth (Brock et al., 1994). Fungi are believed to use their hyphal filaments to physically fragment substrates, encapsulate, and penetrate particles to which the nutrients, or the xenobiotic may be adsorbed (Riser-Roberts, 1992). The lipophilic partitioning characteristics of hydrocarbons to biofilms and organic coatings on benthic substrates, suspended particulates, and dissolved materials will determine its availability to biodegradation and the types of microorganisms performing the degradation (Karickhoff, 1979; Karickhoff et al., 1984; Riser-Roberts, 1992). However, most microorganisms are able to optimize their utilization of any given micro-environment by using multiple function enzymes that are able to shift metabolic pathways to enable the uptake and use of mixed substrates (Brock et al., 1994).

The microbial communities that are specifically responsible for the degradation of hydrocarbons in aquatic ecosystems are heterotrophic bacteria, filamentous fungi, and yeasts or unicellular fungi (Davis, 1967; Riser-Roberts, 1992). Complete mineralization of most complex mixtures of hydrocarbons requires the synergistic associations between both the fungal and bacterial populations (Riser-Roberts, 1992).

The heterotrophic bacteria are the predominant microorganisms involved in the oxidative degradation, assimilation, and cometabolism of hydrocarbons. They possess active mixed function oxidases that are capable of utilizing molecular oxygen to initiate the degradative process. They preferentially utilize the low molecular weight, slightly water soluble, and weakly adsorbed hydrocarbons that include the short chain alkanes and the alkyl-substituted mono-aromatics (Riser-Roberts, 1992). Their small size, short generation times, and capacity to colonize and utilize these hydrocarbons rapidly, enables them to compete more successfully than the fungi for readily available substrates (Davis, 1967; Riser-Roberts, 1992). In environments subject to turbulence from air currents, wave action, water currents, tidal influences, and anthropogenic activities, bacteria are the dominant hydrocarbon degraders (Atlas and Bartha, 1993).

Filamentous fungi are predominantly involved in the oxidative and hydroxylative degradation of the more refractory hydrocarbons. These hydrocarbons have high

molecular weights, are chemically complex in structure, insoluble in water, and have high lipophilic and adsorptive properties. These include the longer chain alkanes, the higher molecular weight substituted aromatics, and the polycyclic aromatics (Riser-Roberts, 1992). Fungi have nonspecific enzyme systems that allow them to degrade or transform hydrocarbons of complex structure, but their metabolism often results in persistent intermediates that are carcinogenic (Riser-Roberts, 1992). Filamentous fungi are slow growing, forming mats or clumps of hyphal filaments that penetrate and fragment the substrates, exposing more surface area to potential degradation (Atlas and Bartha, 1993). In undisturbed environments, fungi can be the dominate microorganism that initiates the degradative process by the physical penetration of the organic pollutant in the surface microlayer (Riser-Roberts, 1992).

The process of hydrocarbon biodegradation is defined as the ability to convert hydrocarbons to compounds of lower molecular weight by the removal of two carbons, as carbon dioxide, through microbially mediated enzymatic oxidation or hydroxylation reactions. The types of hydrocarbons that are actively biodegraded include the alkanes, branched alkanes, alkyl-substituted side chains on cyclic and aromatic hydrocarbon ring structures, alkenes, cycloalkanes, and aromatics (Figures 1-2) (Johnson, 1964). Mineralization is defined as the complete oxidative or hydroxylative degradation of the hydrocarbon to its inorganic components of carbon dioxide and water (Swindoll et al., 1989).

Cometabolism or cooxidation is defined as the simultaneous oxidation of a hydrocarbon by a bacterial microorganism that is actively oxidizing a different hydrocarbon to use as a carbon substrate for growth (Alexander, 1985; Atlas, 1978; Gibson, 1977, 1978; Horvath, 1972; Horvath and Alexander, 1970; Perry, 1979). The oxidizing enzyme recognizes both the substrate for utilization and the other, proximal, hydrocarbon and oxidizes both at the same time. The oxidation process is limited by the greater specificity of the next enzyme in the sequence, which does not recognize the cooxidized hydrocarbon as a substrate. The original substrate is oxidized further, but not the cooxidized hydrocarbon. In some instances, the intermediates can accumulate to very high concentrations that may be inhibitory or toxic to the microbial communities (Atlas and Bartha, 1973; Kappeler and Wuhrmann, 1978a). However, most intermediates are utilized rapidly by the microorganisms having the appropriate enzymatic systems. Prior to utilization, all intermediates must be present in a sufficient threshold concentration to induce the mixed function oxidases in the microorganisms to continue the oxidation process (Atlas and Bartha, 1993; Horvath and Alexander, 1970; Perry, 1979).

Single populations of microorganisms are capable of initiating the oxidation of hydrocarbon components, but complete mineralization and cometabolic processes can only be accomplished by consortia of diverse microbial populations (Horvath, 1972; Horvath and Alexander, 1970; Perry, 1979; Riser-Roberts, 1992). The inability of early researchers to demonstrate microbial degradation of certain hydrocarbons was not necessarily based on the microorganisms' inability to oxidize or cometabolize the hydrocarbons. The experimental methodology that was used, preferentially selected for specific microorganisms that were only capable of utilizing one hydrocarbon as their sole source of carbon and energy (Horvath, 1972).

Early studies of microbially-mediated hydrocarbon degradation mechanisms and rates focused on using laboratory cultured, pure strains of microorganisms, either individually or in artificially assembled consortia. Nutrient amended media containing the individual hydrocarbon of interest would be placed in flasks as a liquid broth or combined with agar to make solid media plates. The selected microorganism(s) were inoculated into the flasked solutions, or plated on the solid medium and then incubated (Atlas and Bartha, 1993; Foster, 1962; Johnson, 1964). The ability of the organisms to grow in the hydrocarbon amended media was considered as evidence of the utilization of the compound as a growth substrate. Quantification of the microorganisms were by plate counts and most probable number. Plate counts consisted of counting the discrete colonies that were formed from the growth of a single microorganism originally deposited on the agar plate. Most probable number uses statistical analyses and successive dilutions of the sample to reach a point of extinction. Replicate dilutions were scored as positive or negative and the pattern of the scores, used with the appropriate statistical tables, gave the most probable number (Atlas and Bartha, 1993). The most probable number technique is more laborious and less precise than the plate count method.

Identification of the microorganisms were determined using one of three methods. The first two methods were based on the phenotypic characteristics of the microorganisms that require specific organic and inorganic nutrients to grow, use specific pathways for the metabolism of nutrients, or are resistant to certain antibiotics. Adjustments to the culture medium would preferentially select for specific phenotypes and inhibit others. The first method consisted of using a selective medium that specifically inhibited the growth of one species, but allowed the growth of another. The second method involved a differential medium amended with specific dyes that indicated by the color produced, the specific metabolic pathway utilized by the microorganism. The third method involved the visual microscopic inspection of the

microorganisms for distinguishing features, such as pigmentation (Atlas and Bartha, 1993; Brock et al., 1994; Kester and Foster, 1963; Pirnik et al., 1974).

Evidence of the biodegradation of the hydrocarbon components were based on the same techniques used to identify and count the microorganisms. Some indicators used were the ability of the microorganism to grow in the hydrocarbon amended medium and the ability to increase in comparison to other microorganisms by presumably utilizing the additional nutrients associated with the hydrocarbons (Horvath, 1972). Another technique that utilized analytical instrumentation involved measuring the aromatic hydrocarbon medium inoculated with microorganisms for ultraviolet properties. The disappearance of ultraviolet absorbance indicated that oxidative cleavage of the aromatic nucleus had occurred and was due to microbially-mediated degradative processes (Atlas, 1978).

These biodegradation tests provided information at the molecular level of the biochemical oxidative mechanisms and transformation products of individual hydrocarbon degradative reactions (Gibson, 1977). The use of radiolabeled oxygen (O^{18}) and carbon (C^{14}) in pure cultures of microorganisms also helped to reveal the mechanisms in the oxidative degradation of *n*-alkanes, branched alkanes, alkenes, cyclics, and some aromatics (Gibson et al., 1970; Gibson et al., 1974; Ooyama and Foster, 1965). The identification of the radiolabeled intermediate compounds and metabolic by-products that accumulated during the degradation process enabled the biochemical sequences to be determined. As analytical methodologies improved, these metabolic by-products were verified by the solvent extraction and analysis of the intermediates using thin layer chromatography, paper chromatography, and partition column chromatography (Johnson, 1964; Kester and Foster, 1963; Pirnik et al., 1974). These results helped to confirm that the primary mechanism of hydrocarbon component degradation was almost exclusively oxidation, performed by specific microbial populations, by means of induced mixed function oxidases utilizing molecular oxygen (Johnson, 1964; Kester and Foster, 1963; Pirnik et al., 1974).

The limitations of the early biodegradation tests using pure microbial cultures in solutions containing individual hydrocarbon became apparent when attempts were made to extrapolate and apply their results to the environment. In the structural and functional complexity of ecosystem-level processes, degradation rate reactions and mechanisms involving complex mixtures of hydrocarbons were inconsistent with laboratory results. The problem was not that the laboratory data were incorrect, but rather that the limited focus of the original experiments were inapplicable to the complexity of interactions and relationships at the ecosystem-level of organization.

The analytical methodologies used in the early microbial studies were based on established enumeration, identification, and interpretive procedures that had been used successfully in the other fields of science during that time. However, the techniques used were inadequate and inappropriate for the small size, short life cycle, short regeneration time, numerical abundance, genetic adaptability, and interactive capabilities specific to microorganisms (Saunders, 1977). The problems associated with enumerating microorganisms using plate counts or most probable number are susceptible to analytical errors as well as methodological errors. Colonies formed too close together may obscure the actual number of originally plated bacteria, while the development of too few colonies will fail to meet the statistical criteria for utilization (Atlas and Bartha, 1993). Nutrient amended agar or media also may contain inhibitory substances that can prevent the full potential for growth and utilization by the microorganisms. Enumerating micron-sized organisms that even at low densities may number in the millions per ml volume is subject to a certain degree of variance and error that is dependent on the analytical technique used and the expertise of the analyst. In addition, the inability of this method to differentiate between viable microorganisms and non-viable, inactive organisms, together with the other factors described, will compromise this methodology in terms of precision, accuracy, and ability to derive meaningful relationships or patterns (Saunders, 1977).

The ability of microorganisms to form cometabolic relationships also is ignored or overlooked as a factor in the biodegradation of the hydrocarbons. The cometabolic process does not yield carbon, energy, or growth to the microorganisms involved. Counts of these microorganisms would yield low population densities in comparison to the quantity and types of hydrocarbons oxidized to their respective intermediate configurations (Alexander, 1985; Horvath and Alexander, 1970).

The use of species identification and classification is another parameter that loses applicability in studies involving microbial organisms (Atlas and Bartha, 1993). Microorganisms are able to alter their DNA and RNA structures by utilizing genetic transfer and recombination mechanisms. These mechanisms in combination with their very short generation times allow them to modify their enzymatic pathways in response to a stressor event or chemical toxicant and adapt to the altered environment. They can then transfer the adaptive mechanism to their progeny and to other microorganisms. The alteration of their genotype will alter their phenotypic utilization of substrates and metabolic pathways. A taxonomically identified microbial population that is exposed to a xenobiotic at the initiation of an experiment will become a different, genetically unique

population by the end of the experiment (Atlas and Bartha, 1993; Ford, 1993; Marshall, 1993).

The application of behavioral patterns observed in larger organisms to explain microbial community interactions and relationships can also lead to incorrect interpretations of metabolic pathways and mechanisms of biodegradation. Competition infers that one species will increase in numbers at the expense of the other species. However, in cometabolic processes, two or more populations could be present that may or may not be numerically the same, based on their hydrocarbon substrate preferences, utilization rates, intrinsic growth rates, and inherent metabolic rates (Horvath, 1972). The diversity of microhabitats available to micron-sized organisms enables these organisms to coexist on temporal and spatial scales that are comparable to no other group of organisms. The microbial relationships and interactions may involve the classical competition, predation, synergistic or antagonistic dynamics observed in more complex organisms, but it is questionable whether they define and dictate microbial behavior with the same importance and to the same extent.

The early attempts of microbiologists to analyze complex microbial processes by reducing the scope of the experimental design to individual species and individual hydrocarbons failed to reveal important microbial interactions and degradative mechanisms for complex mixtures of hydrocarbons. In addition, the lack of methodological standardizations in the types of microbial studies conducted contributed to the difficulty in the validation or comparison of these test results. Test chambers used were different sizes and types. The composition and concentrations of the nutrient amended media solutions and agar plates varied with the diversity of the microorganisms arbitrarily selected for experimentation. Environmental conditions that the microorganisms were incubated at varied in temperature regimes, photoperiods, oxygen concentrations, and illumination. The need to expand the scope of microbial degradative research to include consortia of microorganisms in test systems that are replicable, reproducible, and standardized became apparent.

Hydrocarbon Degradation

Current studies involving indigenous microbial communities have been able to substantiate some of the earlier results as well as provide new dimensions to the characterization of microbial degradation dynamics of hydrocarbon mixtures (Atlas, 1981, 1988, 1991; Evans et al., 1991; Gibson, 1978; Hutchins, 1991; Kroer and Coffin, 1992; Worsey and Williams, 1975). Each chemical class of hydrocarbons including the alkanes, cycloalkanes, alkenes, aromatics, and polycyclic aromatics

(Figures 1-2) are degraded by specific types of microorganism using specific enzymatic mechanisms (Table 1)(Appendix A) (Atlas, 1981; Cerniglia et al., 1979; Gibson, 1977; Focht et al., 1990; Johnson, 1964; Kester and Foster, 1963; Walker et al., 1975; Westlake et al., 1974). The most important microbial species that are responsible for the oxidative degradation of petroleum hydrocarbons belong to the genera *Arthrobacter* and *Pseudomonas* (Atlas, 1981). Once the hydrocarbon is initially oxidized by a specific microorganism, the degradative mechanisms shift from hydrocarbon metabolism to conventional biochemical pathways. The intermediate hydrocarbon is then metabolized as an alcohol, aldehyde or fatty acid, without regard to the origin of the parent hydrocarbon (Figure 3) (Ooyama and Foster, 1965).

Microbial degradation of alkanes is primarily by the sequential oxidation of the terminal methyl group by molecular oxygen by the induction of microbial mixed function oxidases. The intermediates that are formed are more polar compounds, such as alcohols, aldehydes, and fatty acids (Brock et al., 1994; Riser-Roberts, 1992). The initial step in the oxidation of the alkane is believed to start by the attachment of the microorganism to the alkane. The terminal methyl group becomes part of the phospholipid micelle of the cell membrane that forms a pathway from outside the cell membrane to the site of enzymatic activity (Johnson, 1964; Perry, 1979). The initial oxidative attack forms an alcohol that through repeated oxidations produces an aldehyde and finally a fatty acid. The fatty acid can be oxidized at the carbon atom that is *beta* to the initially oxidized carbon atom and the two carbons cleaved off as carbon dioxide to form a new fatty acid, two carbon units shorter. The alkane is then ready for another oxidative attack at the new terminal carbon. This processes continues until acetyl-coenzyme A is the final product, where it is incorporated into the tricarboxylic acid cycle and converted to carbon dioxide (Figure 3) (Atlas and Bartha, 1993; Brock et al., 1994; Davies and Hughes, 1968; Johnson, 1964; Manahan, 1989). Some fatty acid intermediates have been found to be toxic and to accumulate during biodegradation (Atlas and Bartha, 1973). Other fatty acids are incorporated directly into the membrane lipids without further *beta*-oxidation (Riser-Roberts, 1992).

The alkane chains containing five to eighteen carbon atoms are preferentially used as growth substrates though carbon chains with as few as two, or as many as forty-four, can be utilized by some microorganisms (Fredericks, 1966; Riser-Roberts, 1992). Alkanes with shorter chains from five to nine carbon atoms are more easily used as a source of carbon than longer chains with ten to fourteen carbons (Fredericks, 1966). The lower molecular weight alkanes from two to four to are more toxic to bacteria due to their ability to solubilize lipid membranes (Fredericks, 1966).

Table 1. Some microbial species known to be responsible for the oxidative degradation of petroleum hydrocarbons and their metabolic by-products.

Hydrocarbon	Microorganism	Type	Product	Reference
Alkanes	*Pseudomonas* sp.	bacteria-rod	carboxylic acids	Riser-Roberts, 1992
	Norcardia sp.	bacteria-filamentous		Gibson et al, 1974
C_3 - C_6	*Mycobacterium smegmatis*	bacteria-rod	ketones, aldehydes	Riser-Roberts, 1992
			2-butanone	Riser-Roberts, 1992
			ketones, alcohols	Riser-Roberts, 1992
C_{12} - C_{13}	*Cornebacterium* sp.	bacteria club-shaped	fatty acids, alcohols	
	Pseudomonas aeruginosa	bacteria-rod	ketones, esters,	Hou, 1982
			tridecane-1-ol	Kester and Foster, 1963
Tetradecane	*Micrococcus cerificans*	blue-green alga-cocci	myristyl palmitate	Van Eyk and Bartels, 1968
	Cornebacterium sp.	bacteria club-shaped	ketones, esters of	Riser-Roberts, 1992
C_{14} - C_{16}	*Cornebacterium* sp.	bacteria club-shaped	aliphatic acids	Kester and Foster, 1963
Branched	*Brevibacterium erythrogenes*	bacteria coccus-rod	succinyl-CoA	
	Cornebacterium sp.	bacteria club-shaped	succinyl-CoA	Riser-Roberts, 1992
Alkenes	*Pseudomonas oleovorans*	bacteria-rod	1,2-epoxides	Riser-Roberts, 1992
Aromatics	*Achromobacter* sp.	bacteria-rod	3-methylcatechol	Rochkind et al, 1986
	Comamonas testosteroni			Rochkind et al, 1986
Toluene, Kerosene	*Norcardia* sp.	bacteria-filamentous	o-Toluic acid	Kampfer et al, 1991
Toluene, Xylenes,	*Pseudomonas* sp	bacteria-rod	catechols, benzoate	Rochkind et al, 1986
	Cornebacterium sp.	bacteria club-shaped		Gibson et al, 1974
Toluene	*Vibrio* sp., *E. coli*	anaerobic bacteria-rod		Fedorak and Westlake, 1981
	Acinetobacter	bacteria-rod		Fedorak and Westlake, 1981
	Brevibacterium	bacteria-rod		Fedorak and Westlake, 1981
	Flavobacterium	bacteria-rod		Fedorak and Westlake, 1981
Xylenes	*Candida*	fungi-unicellular	o,p-Toluic acid	Fedorak and Westlake, 1981
	Micrococcus	blue-green alga-cocci		Riser-Roberts, 1992
Benzene	*Alcaligenes* sp., *Moraxella*	bacteria-rod, cocci		Fedorak and Westlake, 1981

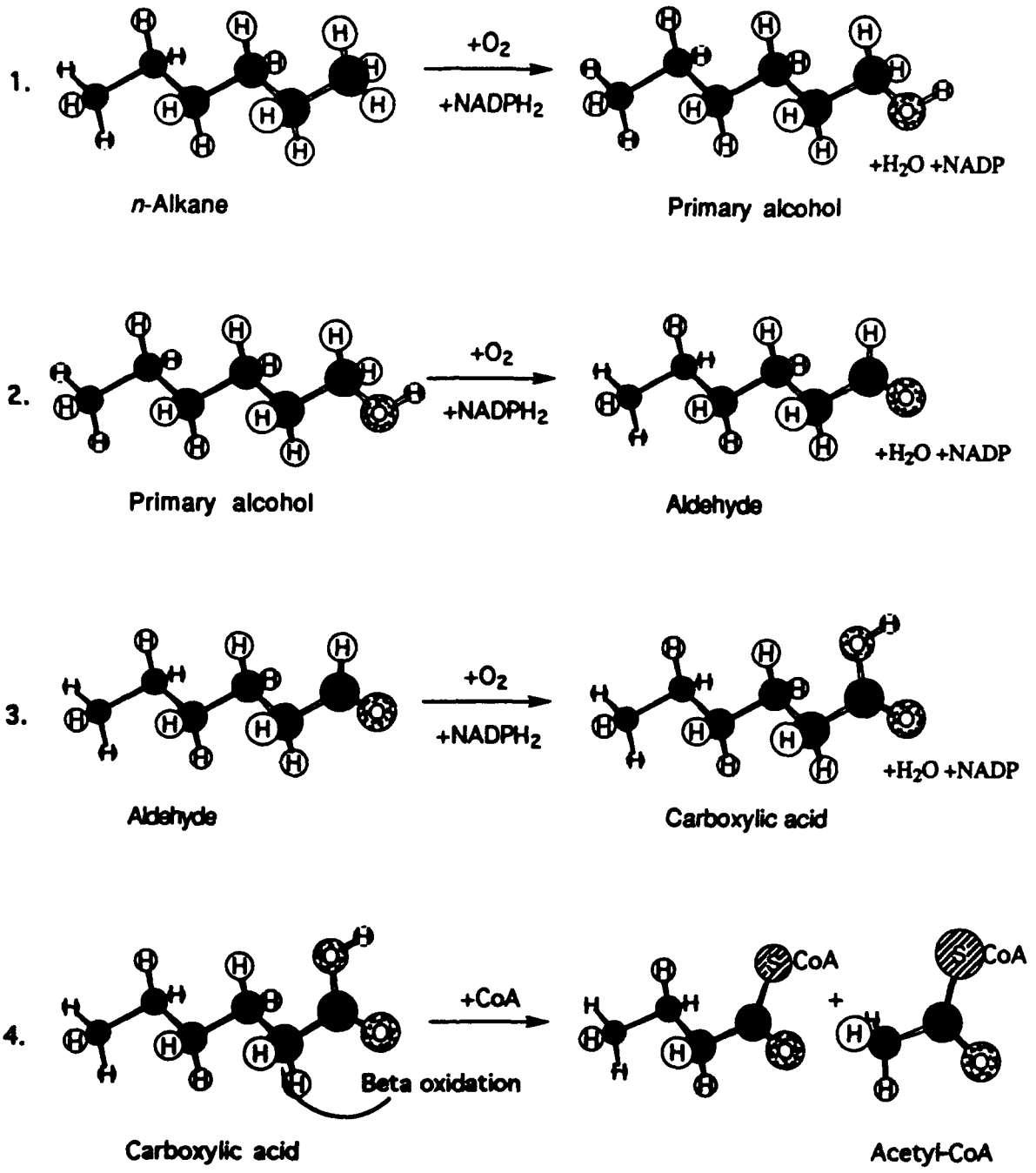


Figure 3. Microbial metabolic pathway for the oxidation of *n*-alkanes and alkyl side chains on cyclic and aromatic compounds (Atlas and Bartha, 1993).

Alkane degradation is predominantly by bacterial populations that can out-compete slower growing microorganisms due to their rapid growth rates and ability to use these easily degraded compounds as a nutritive source. The slower growing microorganisms are able to utilize the more resistant branched alkanes, cyclics, and aromatics (Johnson, 1964; Riser-Roberts, 1992; Schaeffer et al., 1979). Once the alkanes are depleted the bacteria are replaced by the slower growing microorganisms as well as the fungi that have greater metabolic flexibility to degrade refractory hydrocarbons (Fredericks, 1966).

Branched alkanes were found to be similarly degraded, but via the oxidation of both terminal methyl groups (Kester and Foster, 1963; McKenna, 1977; Pirnik et al., 1974). The degree of branching, the location of the side chain, and the length of the parent alkane chain will determine the rate of degradation. Usually, the greater the degree of branching the slower the rate of degradation. Alkanes composed of five to six carbons atoms having an alkyl side chain on the *beta* carbon atom will not be utilized as rapidly as alkanes with seven to ten carbons (Johnson, 1964; Schaeffer et al., 1979). Few microorganisms are capable of utilizing these hydrocarbons for growth due to the limitations of the oxidative enzymes to accommodate the alkyl-branched structure (Riser-Roberts, 1992).

Cycloalkanes are more resistant to degradation than alkanes or branched alkanes and are more toxic. The presence of an alkyl side chain will promote oxidation by providing a terminal methyl group for oxidative attack, followed by the oxidation of the ring structure to form a cyclo-ketone or cyclo-alcohol (Fredericks, 1966). The ring structure is then cleaved and degraded as an alkane (Figure 4) (Atlas, 1981; Ooyama and Foster, 1965).

Alkenes can be oxidized at the unsaturated terminal carbon by the same mechanisms involved in the oxidation of alkanes. They can be oxidized also at the double bond to form epoxides that can be further oxidized to ketones, aldehydes, or esters dependent on the position of the double bond in the chemical structure (Atlas and Bartha, 1993). The 2-alkenes are degraded more readily than the 1-alkenes due to the terminal methyl groups at each end of the molecule that provide sites for potential oxidative attack by more microorganisms (Riser-Roberts, 1992).

Aromatic hydrocarbons are degraded by microbial populations utilizing the same enzymatic pathways, regardless of whether the process occurs under aerobic or anaerobic conditions (Figure 4). Oxygen is the primary electron acceptor in oxic conditions while nitrate, sulfate, carbon dioxide, or nitrous oxide serve as the terminal electron acceptor in anoxic conditions (Atlas, 1981, 1991; Focht, 1988; Hutchins,

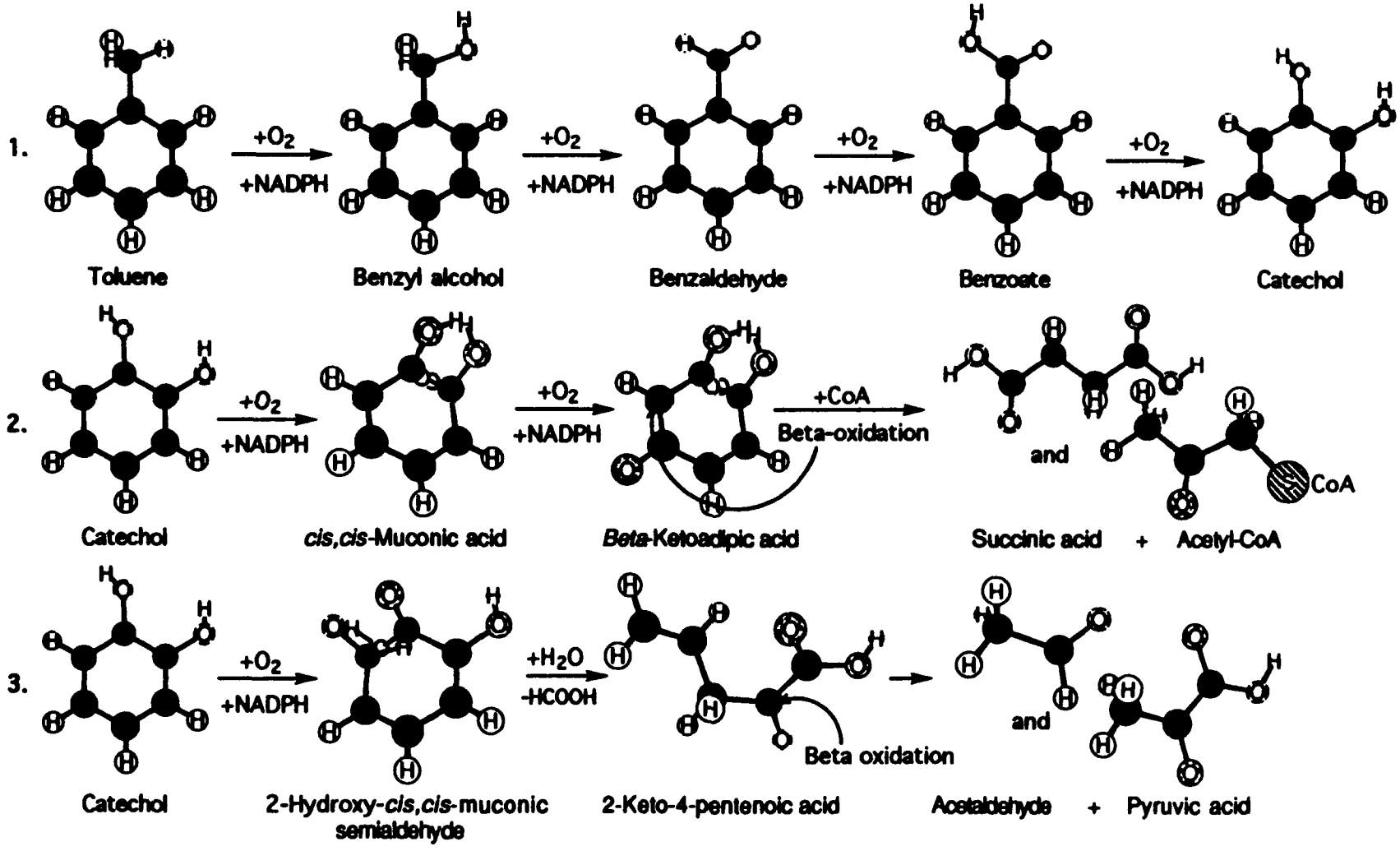


Figure 4. Microbial metabolic pathway for the oxidation of aromatic compounds by *ortho*, Step 2, or *meta*, Step 3, cleavage of the aromatic ring (Atlas and Bartha, 1993; Worsey and Williams, 1975).

1991; Swindoll et al., 1989). The induced microbial mixed function oxidases cause the ring structure to incorporate one molecule of oxygen in the form of a dioxygenase and the resulting dihydroxylated ring (catechol) is opened by enzymatic cleavage at the *ortho* position, yielding a *cis,cis*-muconic acid. This is oxidized further to *beta*-ketoadipic acid which is oxidatively cleaved into the tricarboxylic cycle intermediates of succinic acid and acetyl-Coenzyme A (Atlas and Bartha, 1993; Focht, 1988; Gibson, 1977).

A alternative metabolic pathway involves the cleavage of the dihydroxylated aromatic ring at the *meta* position to yield 2-hydroxy-*cis,cis*-muconic semialdehyde that after further metabolism produces 2-keto-4-pentenoic acid, acetaldehyde, and pyruvic acid (Atlas and Bartha, 1993; Rochkind, 1986). Fungi and algae are able to biodegrade aromatic hydrocarbons similar to the mechanism utilized by mammalian hepatic mixed function oxidases where the *trans*-benzene-1,2-oxide intermediate is formed rather than the *cis*-benzene-1,2-oxide formed by bacteria (Cerniglia et al., 1979; Gibson, 1977). The same metabolic pathways are also involved in the degradation of polycyclic hydrocarbons (Davies and Hughes, 1968). The *trans*-diol intermediates formed in the degradation of many polycyclic aromatic hydrocarbons have been found to be carcinogenic while *cis*-diols have not (Atlas and Bartha, 1993).

The degradation of alkyl-substituted aromatics is initiated first by the oxidation of the terminal methyl group followed by *beta* oxidation, as in the oxidation of alkanes. The ring is then oxidized and cleaved. The presence of an methyl or alkyl side-chain on the aromatic ring will increase the rate of oxidative degradation. In addition, mono-aromatics will be degraded faster than the di-, tri-, and poly-cyclic aromatics (Atlas, 1978, 1981; Cerniglia et al., 1979; Gibson, 1977).

The specific combination of hydrocarbon components in the fuel mixtures will determine the types of microbial populations actively present and the degradation rates of the individual hydrocarbons (Bailey et al., 1973; Horowitz and Atlas, 1977b; Ooyama and Foster, 1965; Perry, 1979; Walker and Colwell, 1976a; Walker et al., 1976b, 1976c; Westlake et al., 1974). In addition, the molecular weight, structural configuration, and concentration, of the hydrocarbon component will mediate the rate of oxidation. The lower the molecular weight and the simpler the hydrocarbon structure the more rapid will be the degradation (Atlas and Bartha, 1993; Davis, 1967). The rank order of hydrocarbon utilization and degradation rates from the most rapid to the slowest is: alkanes, branched alkanes, cyclo-alkanes, aromatics, polycyclic aromatics, and asphaltenes. This relationship does not imply that these compounds are degraded sequentially. The lower molecular weight fractions of all hydrocarbon classes are degraded at the same time, but a different rates that are dependent on their concentration

in the mixtures (Bailey et al., 1973; Horowitz and Atlas, 1977b; Ooyama and Foster, 1965; Perry, 1979; Walker and Colwell, 1976a; Walker et al., 1976c; Westlake et al., 1974).

In a mixture containing high concentrations of *n*-alkanes Walker, Colwell, and Petrakis (1976c) found that the alkanes were degraded more rapidly than aromatics. In a mixture containing high levels of aromatics the same microbial communities degraded more aromatics than alkanes. However, each mixture supported the growth of different populations of microorganisms with the *n*-alkane mixture supporting bacteria and yeasts and the aromatic mixture supporting only bacteria (Walker et al., 1975; Walker and Colwell, 1976a; Westlake et al., 1974). The microorganisms that were not utilizing a specific hydrocarbon substrate for growth were still found to be present, but in a dormant or inactive state.

The higher molecular weight aromatic hydrocarbons are degraded at the same time as the lighter hydrocarbons, but their structural complexity makes these compounds more resistant to oxidative attack and subsequently slower to be degraded. The presumed preferential degradation of lower molecular weight hydrocarbons by microorganisms is a misconception. The presence of two potential sites for oxidative attack at the terminal methyl groups of straight chain alkanes and the ease of the oxidation determines the faster rate of degradation for these compounds (Bailey et al., 1973). Sequential degradation of hydrocarbon components would result in a lag period in the biodegradation of the other classes of hydrocarbon components (Atlas and Bartha, 1972; Walker and Colwell, 1976a). Gas chromatograms of the water soluble hydrocarbon components of jet fuel display no lag periods in the degradation of the higher molecular weight compounds (Figures 7-12).

The cometabolic oxidation of many hydrocarbons will be determined by the specific hydrocarbon components present in the petroleum mixture (Horvath, 1972; Horvath and Alexander, 1970; Ooyama and Foster; 1965; Perry, 1979). Aromatic hydrocarbons were initially believed to be resistant to microbial oxidations due to the inability of single species of microorganisms to oxidize and utilize the hydrocarbons as nutritive substrates (Horvath, 1972). However, when the aromatics are present as mixtures with other hydrocarbons that can serve as nutritive substrates, a substantial number of the mono-, di-, and tri-aromatic hydrocarbons can be cometabolized by consortia of microorganisms (Alexander, 1980; Atlas, 1978; Gibson, 1977; Horvath, 1972; Horvath and Alexander, 1970). Cometabolism of other hydrocarbons were also found to require specific hydrocarbon substrates. Van Eyk and Bartels (1968) found that hexane and butane are strong inducers of alkane degradation. Hexane can serve as a

growth substrate in the cooxidation of 2-methylheptane, *o*-xylene, and ethylcyclohexane. The cometabolism of cycloalkanes requires the presence of propane, *n*-heptane, or 2-methylbutane to serve as substrates (Hou, 1982; Ooyama and Foster, 1965). In the cometabolism of several aromatic hydrocarbons, hexadecane was found to serve as the primary substrate (Perry, 1979).

Several other factors will affect microbial utilization rates. The presence of thiols, metals, and other contaminants in the hydrocarbon mixture can inhibit or be potentially toxic to the microorganisms. The production and secretion of emulsifiers by microorganisms will potentially expose more of the surface area of the hydrocarbon mixture to microbial oxidative activity. The viscosity of the hydrocarbon mixture will determine the dispersal and size of the contaminated area, with the thinner and less viscous layers being more conducive to microbial attack. Finally, the presence, thickness, and organic content of the microlayer will determine the amount of hydrocarbon compounds absorbed and adsorbed as well as the availability of microenvironments and substrates for microbial utilization (Atlas, 1981, 1988, 1991; Focht, 1988; Kampfer et al., 1991; Wolfe, 1987). Microorganisms have been found to occur from ten to one hundred times more frequently in the surface microlayer than at the ten centimeter depth at oiled sites (Atlas, 1981).

Environmental factors are crucial in the determination of the specific indigenous microbial species present and their rates of hydrocarbon degradation. One of the primary factors mediating microbial oxidative degradation is temperature (Atlas et al., 1978; Brock et al., 1994). Temperatures of 20°C to 40°C will increase microbial utilization rates of hydrocarbons. Abiotic losses of lower molecular weight hydrocarbons will also be increased by evaporation and volatilization (Atlas, 1975; Dibble and Bartha, 1979; Horowitz and Atlas, 1977b; Ward and Brock, 1976; Westlake et al., 1974). At lower temperatures near 4°C, growth rates and utilization rates of psychrophilic microorganisms have been found to be similar to mesophilic microorganisms at 18°C (Delille and Siron, 1993). Abiotic losses are dramatically reduced, especially in the presence of ice that will physically prevent the volatilization of hydrocarbons from the water (Atlas, 1975).

Atlas (1975) found that both paraffins and aromatics are biodegraded at 10°C and 20°C in all oils with the lighter alkanes being degraded more rapidly than any of the other types of hydrocarbons. Heavier oils were found to be degraded at slightly higher rates at lower temperatures due to the lack of inhibitory effects to the microbial communities, that is present at higher temperatures. Cometabolism was found to be the primary degradative mechanism at lower temperatures (Horowitz and Atlas, 1977b).

Studies of temperate lakes in Wisconsin by Ward and Brock (1976) found that summer temperatures of 20°C to 25°C were optimal for oil biodegradation by indigenous microbes, but nutrient limitations restricted the rates and extent of the degradation. During the spring, winter, and fall seasons temperature became the primary limiting factor. The highest rates of hydrocarbon degradation were found to be about one month after ice melt, when the constraints imposed by both temperature and nutrient concentrations on microbial utilization rates were minimized.

All environmental factors, when they are examined separately, will impose some constraints on the types of microorganisms present and their functional processes. Each factor will be described in terms of a range that will be optimal for non-specific microbial growth and utilization rates. Some of these ranges have been determined for optimizing microbial utilization rates of hydrocarbons. The optimal range for pH is from 7.5 to 7.8 (Dibble and Bartha, 1979). Oxygen concentrations range from 5 mg/L to 8 mg/L (Ward and Brock, 1976). Nutrient concentrations of nitrogen, phosphorus, and iron must be present in specific proportions to each other with C:N, C:P, and N:P ratios of 10:1, 100:1, and 10:1 to be optimal (Atlas, 1981; Dibble and Bartha, 1979; Fedorak and Westlake, 1981). In addition, other factors including season, climate, sunlight, altitude, depth, wind and wave activity, and prior exposure of the environment to hydrocarbon releases will also affect the microbial community structure and degradation rates (Atlas, 1981, 1987; 1991; Evans et al., 1991; Hutchins, 1991; Focht, 1988; Riser-Roberts, 1992; Wolfe, 1987). Some abiotic rate-dependent processes that will also be affected by these environmental factors include photooxidative decomposition, adsorption, absorption, hydrocarbon partition coefficients, and solubilization.

The results of some of the studies on the effects of environmental factors on hydrocarbon degradation rates appear to be intuitively self-explanatory. Warmer seasons will increase degradation rates, wind and wave action will increase volatilization and mixing to provide greater surface area for degradative activity, and aerobic utilization rates will be faster than anaerobic rates. The limitations of these studies involving single environmental factors are their inapplicability to real environmental conditions. The importance of the study by Ward and Brock (1976) was that optimal conditions for microbial populations are never realized in terms of any single environmental factor. There will always be other factors that will become limiting or impose constraints. Environmental factors when combined, produce a unique set of conditions on temporal and spatial scales that will affect microbial functions and processes in ways that on first examination may be counterintuitive.

The ability to examine microorganisms in the context of ecosystem-level physical, structural, and functional interactions will reveal microbial functional processes as well as the mechanisms that control and regulate these functions. Laboratory experiments and field validation studies must be well defined and designed to be conducive to examining the special functions, processes, and interrelationships that microorganisms impart to all ecosystems. The small size, enzymatic adaptability, short generation time, and genetic versatility of microorganisms conveys to them a uniqueness apart from all other types of organisms. Microorganisms are the single, primary organisms responsible for the fixation of nitrogen, decomposition of organic matter, transformation of chemical contaminants, recycling of nutrients and carbon, and energy flows in all environments. To ignore their importance and role in any study involving ecosystem structural components and functional processes will yield results that are limited in scope, applicability, and ecological significance.

The objectives of this study were to measure the hydrocarbon degradative responses of two separate microbial communities exposed to complex mixtures of hydrocarbons in two types of microcosms, the Mixed Flask Culture microcosm and the Standardized Aquatic Microcosm. The microcosms were different in the source of the organisms, the numbers of organisms, species diversity, trophic level complexity, sediment organic matter quality, and size. The purpose was first to determine whether the microbial communities display similar patterns of rate responses and intensities in the degradation of hydrocarbon components when treated with the water soluble fraction of a jet fuel. The second purpose was to determine whether the same rate patterns are repeated when the microorganisms were re-treated with a second water soluble fraction of jet fuel. Third, to determine whether the patterns were similar to responses and intensities observed in field studies. Finally, to determine whether microcosms must resemble real ecosystems, as closely as possible, to be valid models of ecosystem dynamics, or whether ecosystem dynamics display similar, universal patterns in rate responses independent of species composition and trophic complexity.

Materials and Methods

Chemicals

All chemicals used in the culture of the organisms for the Standardized Aquatic Microcosm and in the preparation of the microcosm medium, T82MV, were reagent grade or as specified by the ASTM and USEPA protocols (ASTM E1366-91, 91; Shannon and Anderson, 1989). Individual hydrocarbon reference standards that were used to identify and quantify the water soluble components in the jet fuels were purchased from the Alltech Chemical Company (Deerfield, IL) and were A.C.S. spectrophotometric grade (≥99+% purity). These included *n*-pentane, *n*-hexane, *n*-heptane, *n*-octane, *n*-nonane, *n*-decane, *n*-undecane, *n*-dodecane, *n*-tridecane, *n*-tetradecane, *n*-pentadecane, cyclohexane, cycloheptane, cyclooctane, benzene, ethylbenzene, toluene, *ortho*-xylene, *meta*-xylene, and *para*-xylene. The ASTM D3710 Qualitative Calibration Mix and the Qualitative Reference Reformate Standard were purchased from Supelco Chromatography Products (Bellefonte, PA). All standards were prepared in pesticide grade, A.C.S. specification hexane or carbon disulfide, purchased from VWR Scientific (Seattle, WA).

The Jet-A jet fuel used in the Mixed Flask Culture microcosm experiment is a refined aviation fuel used extensively throughout the world in commercially operated aircraft and was provided locally by Fliteline Services of Bellingham, Washington. The JP-8 jet fuel used in the Standardized Aquatic Microcosm experiment is a new, experimental formulation refined and produced by the United States Air Force and was provided by the U. S. Air Force Toxicology Laboratory at Wright-Patterson Air Force Base in Ohio. The samples were collected in two liter fuel cans from in-line quality assurance/quality control valves and were sealed on site. The lot shipment was recorded and the samples were transported to the laboratory.

Glassware

All reagent preparation, measuring, and dispensing of solutions were performed using class A volumetric glassware. The preparation, mixing, and addition of the 100% water soluble fraction of the jet fuel to the microcosms were performed using 1 L separatory funnels and class A graduated cylinders. The graduated cylinders were used to reduce the potential loss of volatiles during the measuring and dispensing process. All glassware used in the culture of the laboratory organisms, in the preparation of solutions, reagents, and media including microcosm containers, samplers, and sample reservoirs were washed in hot soapy water using Labtone®, a non-phosphate dishwashing soap. The glassware was rinsed four to six times in hot water until all

soapy residue was removed, rinsed four times in distilled water, inverted and allowed to dry. The dried glassware was covered with aluminum foil with the dull side in contact with the glass and autoclaved for thirty minutes in a Market Forge Sterilemaster at 15 psi and 250°C.

All microcosm beakers, jars, samplers, and sample reservoirs were filled with the 2N HCl after the regular washing and rinsing process and allowed to soak overnight. The following day the glassware was rinsed ten times in distilled water, with the final rinse remaining in the glassware for another twelve hours to remove any residual acid. The glassware was then rinsed two more times with distilled water, inverted and dried. . When dry the glassware was covered with aluminum foil, dull side in contact with the glass, and autoclaved for 30 minutes.

Chromatography glassware consisted of class A volumetric flasks, pre-cleaned 4 ml, 10 ml, and 40 ml clear glass vials with Teflon®-lined screw caps, gas tight Hamilton® syringes, two 5 ml gas tight Teflon® luer lock Hamilton® syringes, and two 5 ml glass spargers with frits for the Tekmar® Purge and Trap LSC 2000 Concentrator. The syringes, vials, and spargers were purchased from Supelco Chromatography Products (Bellefonte, PA).

Chromatography glassware was cleaned in hot soapy water, using Alconox® powder detergent, rinsed in hot water until all soapy residue was removed, rinsed three times with 10% sulfuric acid (H_2SO_4), rinsed ten times in deionized distilled water (obtained from a Barnstead Nanopure four cartridge system) then inverted and dried at 105°C for four hours.

Preparation of the 100% water soluble fraction of jet fuel

The water soluble fraction (WSF) of Jet-A and JP-8 were prepared in glassware washed in nonphosphate soap, rinsed, the soaked in 2N HCl for at least one hour, rinsed ten times with distilled water, dried, and autoclaved for 30 minutes. The Standardized Aquatic Microcosm medium, T82MV, was substituted as the diluent for the water fraction of the WSF. Separatory funnels were used as mixing chambers to prepare the 100% WSF.

Twenty-five ml of the appropriate jet fuel were added to each 1 L separatory funnel containing 1 L of sterile, fresh T82MV medium and mixed by agitating the separatory funnel contents vigorously for five minutes, slowly releasing built up pressure when necessary, allowing the contents to stand undisturbed for fifteen minutes, and repeating this procedure until a total time of one hour had elapsed. The separatory

funnel and its contents were then allowed to remain undisturbed for twelve hours to maximize the saturation of the T82MV with the water soluble components in the jet fuel.

After twelve hours, the T82MV/100% water soluble fraction of jet fuel mixture was carefully drained from the separatory funnel. The final 100 ml of the WSF in direct contact with the jet fuel layer was left in the separatory funnel to avoid incorporating any fuel emulsion into the final water soluble fraction. The 100% WSF was placed directly into clean, sterilized one liter amber glass bottles and capped with Teflon®-lined screw caps. The 100% WSF was used within twelve hours of preparation.

In the Mixed Flask Culture microcosm experiment, two liters of 100% WSF of JET-A were prepared to ensure a final volume of 1260 ml necessary to treat the eighteen 1 L MFC microcosm beakers. In the Standardized Aquatic Microcosm experiment, four liters of 100% WSF of JP-8 were prepared to provide the necessary 3780 ml to treat the eighteen 3 L SAM jars.

Microcosm Protocols

Treatment Regime

Each microcosm experiment consisted of two consecutively conducted component experiments. The first experiment was conducted as specified in the original protocol using 0%, 1%, 5%, and 15% water soluble fractions of jet fuel as the toxicant. The second, extended experiment consisted of re-treating the 0% WSF reference and 15% WSF treated microcosms with fresh 15% water soluble fractions of jet fuel after sixty-three days had elapsed.

Mixed Flask Culture Microcosm (MFC)

Construction and implementation of the 60-day Mixed Flask Culture (MFC) microcosm experiment was conducted to the specifications described in Shannon and Anderson's (1989) modification of Leffler's (1980) original protocol. Natural assemblages of aquatic organisms were collected from local streams and lakes during January of 1992. The sites included Padden Creek, Squalicum Creek, Whatcom Creek, Baker Creek, Silver Creek, Heron Pond, Lake Whatcom, and Lake Padden. The samples were brought back to the laboratory and placed in a 50 L aquarium containing sterile T82MV medium (ASTM E1366-91). The organisms were then allowed to acclimate for a three month period. Laboratory environmental conditions were maintained at 20° ± 2°C with the light intensity at 80 ± 2 $\mu Em^{-2}s^{-1}$ and a photoperiod of 12 hours light and 12 hours dark. At the end of three months the resulting aquatic community was subsampled, with 50 ml placed into each of the thirty, cleaned and acid washed 1 L beakers. Each

beaker contained 50 g of acid-washed white silica sand, 15 μg $NaHCO_3$, and 900 ml of freshly made, sterile T82MV medium. The beakers were placed uncovered in a Puffer-Hubbard CEC 50LTP Environmental Chamber. The environmental conditions were set to an isothermal day/night temperature of 20° ± 2°C, illumination at 80 ± 2 $\mu Em^{-2}s^{-1}$, and a photoperiod of 12 hours light and 12 hours dark (Figure 5).

The MFC's were allowed to equilibrate for six weeks. Once a week they were cross-inoculated to minimize divergence, re-inoculated to ensure a more uniform distribution of organisms among the beakers, and rotated within the environmental chamber to minimize potential light and temperature variations. All microcosms were removed from the environmental chamber, but kept in their order of position. Each microcosm was gently stirred, a 100 ml aliquot removed using a Standardized Aquatic Microcosm dip sampler described in ASTM E1366-91, and combined together in a sterile, 4 L Erlenmeyer flask. To the flask contents, 300 ml from the stock community aquarium were added, mixed, and redistributed to each microcosm in 110 ml aliquots. The final volume of each microcosm was brought up to 1 L with fresh, sterile T82MV medium to compensate for evaporative losses. The microcosms were then placed back into the environmental chamber in a pattern that would rotate each set of four beakers to a new shelf, and clockwise individually by one microcosm position.

After six weeks the microcosms were examined individually to verify that each contained the specified minimum functional groups: two species of unicellular green algae, one species of nitrogen fixing blue-green algae, one species of filamentous green algae, one species of herbivorous grazer, one species of benthic detritivore, bacteria, and protozoans. The microcosms were then monitored for two days, with morning dissolved oxygen, evening dissolved oxygen, and pH measurements recorded. The microcosms that displayed measurements divergent from the mean and outside the determined 95% Confidence Intervals were removed from the experiment. A final total of twenty-four microcosms were then randomly numbered and assigned to four treatment groups (including the reference group) with each group containing six replicates. The microcosms were re-assigned shelf positions within the environmental chamber with each shelf containing one replicate from each treatment group.

JET-A Jet Fuel Water Soluble Fraction Application

On Friday (designated Day 0) 150 ml of the medium was removed from each of the twenty-four MFC's using an autoclaved, 100 ml capacity basting tube, with a sterile square of 100 mesh Nitex® tied over the opening to prevent the removal of the

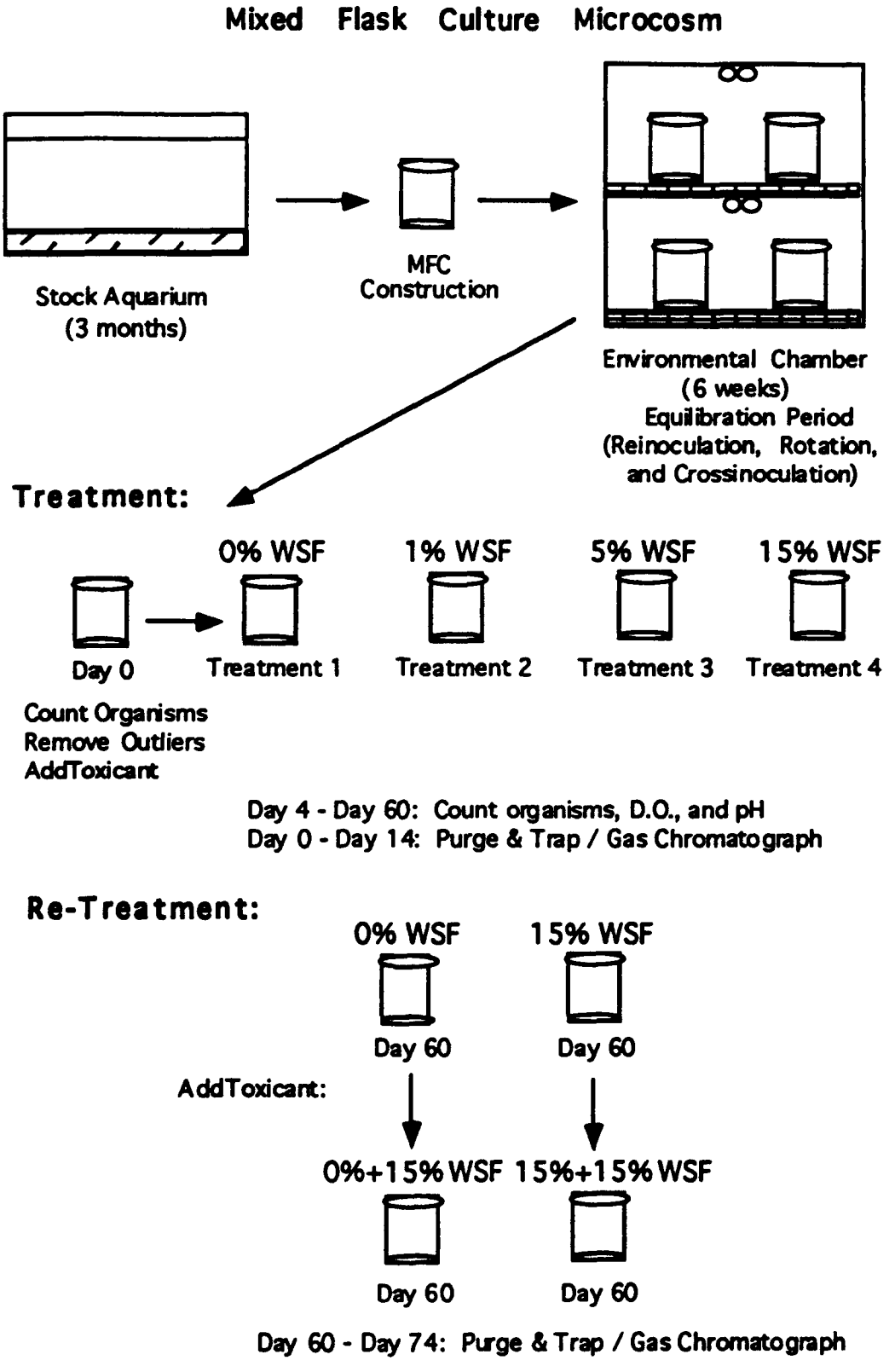


Figure 5. Mixed Flask Culture microcosm construction and JET-A water soluble fraction treatment application.

organisms. Treatment groups (of six replicates each) containing 0%, 1%, 5%, and 15% JET-A water soluble fractions were prepared as follows:

1) 0% WSF (designated as the reference group) were refilled with 150 ml of fresh, sterile T82MV medium;

2) 1% WSF (designated as Treatment 2) were refilled with a 150 ml mixture containing 10 ml of the 100% JET-A water soluble fraction combined with 140 ml of fresh, sterile T82MV medium;

3) 5% WSF (designated Treatment 3) were refilled with a 150 ml mixture containing 50 ml of the 100% JET-A water soluble fraction combined with 110 ml of fresh, sterile T82MV medium;

4) 15% WSF (designated Treatment 4) were refilled with 150 ml of the 100% JET-A water soluble fraction jet fuel.

The final volume in each microcosm was adjusted to 1 L with sterile, fresh T82MV. The beakers were covered with sterile 15 mm x 100 mm inverted plastic petrie dish lids and placed in their assigned positions within the environmental chamber. The microcosms were monitored for physical, structural and functional parameters on Tuesdays and Fridays for the duration of the 60 day experiment. The results from this experiment are presented by Landis et al.,(1994).

On day 60 both the 0% WSF reference group and the 15% WSF treatment group were split into two sub-groups of three replicates. Three replicates from the reference treatment group and three replicates from the 15% WSF treatment group were treated with freshly prepared 100% WSF JET-A jet fuel to a final concentration of 15% WSF (Figure 5). The other six replicates were treated with freshly prepared T82MV medium. As before, 150 ml were removed from each of the twelve microcosms and treatment groups (of three replicates each) containing 0%+0%, 0%+15%, 15%+0%, and 15%+15% JET-A water soluble fractions were prepared as follows:

1) 0%+0% WSF (designated as the new reference group Treatment 1) were refilled with 150 ml of fresh, sterile T82MV medium;

2) 0%+15% WSF (designated as the new Treatment 2) were refilled with 150 ml of freshly prepared 100% WSF of JET-A;

3) 15%+0% WSF (designated as the second reference group Treatment 3) were refilled with 150 ml of fresh, sterile T82MV medium;

4) 15%+15% WSF (designated as the new Treatment 4) were refilled with 150 ml of freshly prepared 100% WSF of JET-A.

The final volume in each microcosm was adjusted to 1 L with sterile, fresh T82MV. The beakers were then covered with sterile 15 mm x 100 mm inverted plastic petrie dish lids and returned to the environmental chamber.

Standardized Aquatic Microcosm (SAM)

The 63-day Standardized Aquatic Microcosm (SAM) protocol is described in ASTM E1366-91. The microcosms are constructed by placing a sediment of 200 g of acid-washed, white silica sand, 0.5 g of cellulose powder, and 0.5 g of ground chitin in the bottom of each of thirty, 3.8 L glass jars and adding 3 L of unautoclaved T82MV medium. The jars were placed in an autoclavable Nalgene® tray that was filled with enough water to be above the level of the sediment in the jars and autoclaved for one hour at 250°C and 15 psi. After cooling, the filter-sterilized vitamins and other heat sensitive T82MV medium components were added to each jar and the pH adjusted to 7.0 ± 0.2 with sterile 1N HCl. The jars were then capped to prevent airborne contamination until the initiation of the experiment, within twenty-four hours after the construction of the microcosms.

At the initiation of the experiment (designated Day 0) nine axenic laboratory cultures of algae were harvested, rinsed with fresh, sterile T82MV medium, counted, and inoculated at the recommended density of three million cells into each microcosm (Figure 6). The species used were *Anabaena cylindrica, Ankistrodesmus* sp., *Chlamydomonas reinhardi* 90, *Chlorella vulgaris, Lyngbya* sp., *Scenedesmus obliquus, Selenastrum capricornutum, Stigeoclonium* sp., and *Ulothrix* sp. The unicellular green algae were counted using a hemacytometer. The filamentous green algae and the filamentous blue-green algae were counted by initially breaking the filaments into shorter strands. The algae were placed in sterile, 50 ml culture tubes with Teflon®-lined screw caps containing acid washed glass beads and agitated vigorously. Nanoplankton counting chambers, or Palmer cells, were then used to count the individual cells.

The SAM microcosms were each covered with a sterile 15 mm x 150 mm inverted plastic petrie dish lid and placed in random positions in an oval pattern on a specifically constructed table in an environmentally controlled room. The bank of lights

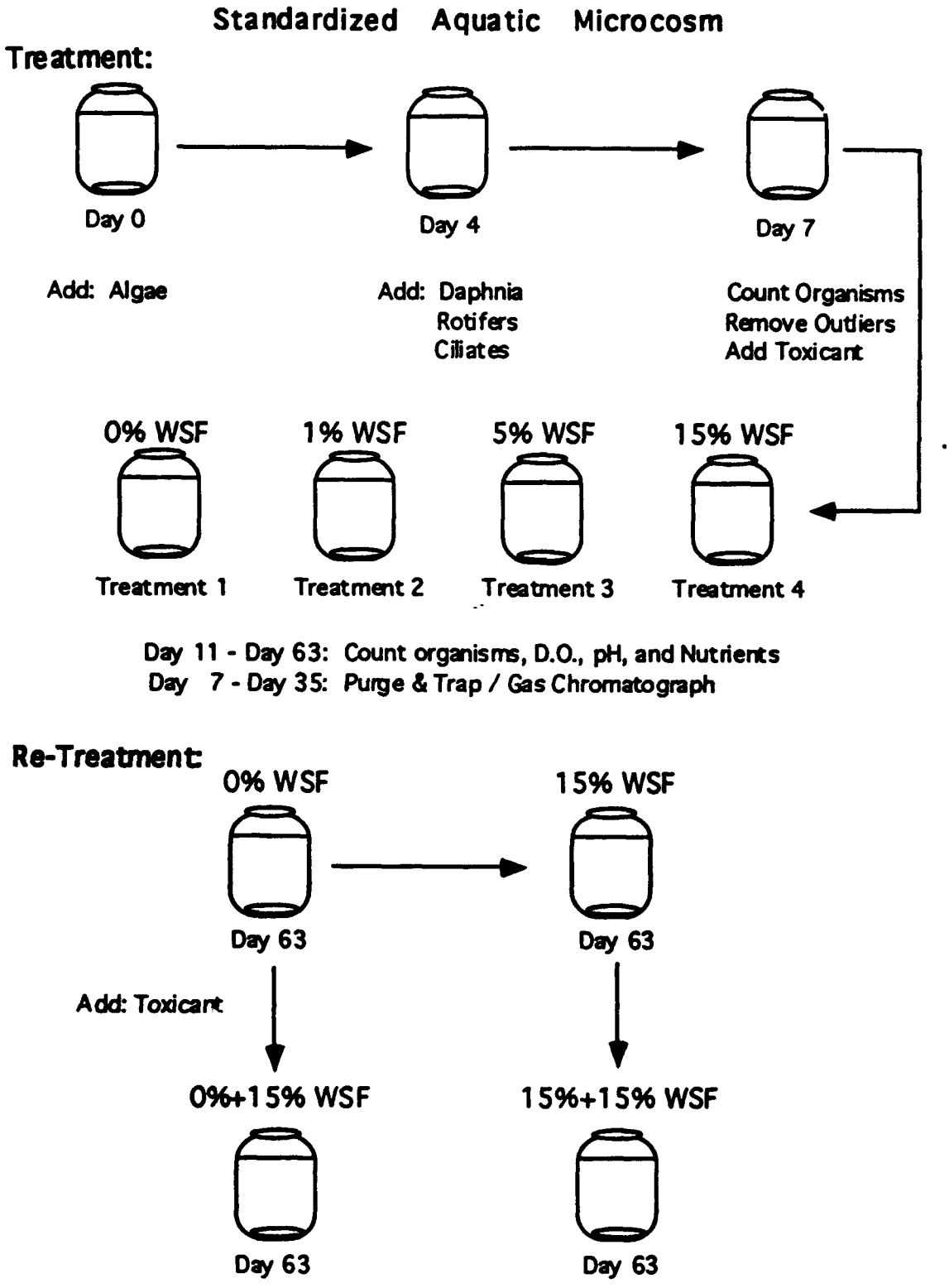


Figure 6. Standardized Aquatic Microcosm construction and JP-8 water soluble fraction treatment application.

parallel above the microcosms were adjusted to provide illumination at 80 ± 2 $\mu Em^{-2}s^{-1}$, with a photoperiod of 12 hours light and 12 hours dark. The temperature in the room was maintained at an isothermal day/night temperature of 20° ± 2°C.

On Day 4 the remaining organisms were added to each microcosm in the following densities: 300 *Tetrahymena* sp., 90 *Philodina* sp., six *Cyprinotus* sp., and 6 adult *Daphnia magna* with eggs, 6 immature *D. magna* without eggs, and 10 *D. magna* instars (less than 24 hours old). The microcosms were then monitored for two days, with morning and evening pH and dissolved oxygen measurements recorded to identify and remove microcosms outside the determined 95% Confidence Intervals for a final total of twenty-four microcosms. The SAM's were then randomly numbered and assigned to four treatment groups, each containing six replicates, with one treatment group of six replicates (Treatment 1) serving as the reference. The replicate microcosms were assigned to six block groups of four around the table so that each block held one replicate from each treatment group.

On Day 7, a 600 ml subsample was removed from each of the twenty-four microcosm and placed in twenty-four correspondingly labeled, sterile 1 L temporary reservoirs. Algae were counted in each reservoir using a Palmer cell with a 10x ocular and 40x objective phase contrast microscope. Twenty-five microscope fields or 50 cells were counted for each algal species. Counts of daphnia and ostracods were made by visually inspecting the entire 600 ml subsample. The microorganisms were counted by dispensing 0.2 ml aliquots of a 2 ml subsample onto a clean petrie dish lid, inspecting each drop using a stereozoom microscope, and recording the total number per 2 ml sample. After all counts were made the 600 ml subsamples were returned to their respective microcosms.

JP-8 Jet Fuel Water Soluble Fraction Application

On Day 7, after all counts were completed and the subsamples returned to their respective microcosms, 450 ml of medium were removed from each of the twenty-four 3 L microcosms using the sterile basting tube with Nitex®. Treatment groups containing 0%, 1%, 5%, and 15% JP-8 water soluble fractions were prepared as follows:

1) 0% WSF (designated as the reference group Treatment 1) were refilled with 450 ml of fresh, sterile T82MV medium;

2) 1% WSF (designated as Treatment 2) were refilled with a 450 ml mixture containing 30 ml of the 100% JP-8 water soluble fraction prepared earlier, combined with 420 ml of fresh, sterile T82MV medium;

3) 5% WSF (designated as Treatment 3) were refilled with a 450 ml mixture containing 150 ml of the 100% JP-8 water soluble fraction combined with 300 ml of fresh, sterile T82MV medium;

4) 15% WSF (designated as Treatment 4) were refilled with 450 ml of the 100% JP-8 water soluble fraction jet fuel.

The final volume in each microcosm was adjusted to 3 L with sterile T82MV medium and the microcosms re-covered with a sterile 15 mm x 150 mm inverted plastic petrie dish lid. The microcosms were monitored for structural parameters, with subsamples removed twice weekly from each microcosm and counts of population densities made for all species for the duration of the 63 day experiment. The results from the SAM microcosm are presented by Landis et al. (1994).

After day 63, both the reference (0% WSF) and the 15% WSF treatment groups were split into two sub-groups of three replicates. Three replicates from the reference treatment group and three replicates from the 15% WSF treatment group were treated with freshly prepared 100% WSF JP-8 jet fuel (Figure 6). The other six replicates were treated with freshly prepared T82MV medium. As before, 450 ml were removed from each of the twelve microcosms and treatment groups (of three replicates each) containing 0%+0%, 0%+15%, 15%+0%, and 15%+15% JP-8 water soluble fractions were prepared as follows:

1) 0%+0% WSF (designated as the new reference group Treatment 1) were refilled with 450 ml of fresh T82MV medium;

2) 0%+15% WSF (designated as the new Treatment 2) were refilled with 450 ml of freshly prepared 100% WSF JP-8;

3) 15%+0% WSF (designated as the second reference group Treatment 3) were refilled with 450 ml of fresh T82MV medium.

4) 15%+15% WSF (designated as the new Treatment 4) were refilled with freshly prepared 100% WSF JP-8.

The final volume in each microcosm was adjusted to 3 L with sterile T82MV medium, re-covered with a sterile 15 mm x 150 mm inverted plastic petrie dish lid and placed in a randomly assigned a new position around the table in the environmentally controlled room.

Sterile Mixed Flask Culture Microcosm

Three additional microcosms that were conditioned for six weeks were autoclaved in a Market Forge autoclave for one hour at 250°C and 15 psi. Once cooled, 150 ml of the medium were replaced with 150 ml of fresh made 100%/T82MV water soluble fraction of JET-A. The sterile MFC's were covered with a sterile 15 mm x 100 mm inverted plastic petrie dish lid and placed in the environmental chamber. Samples were removed and analyzed on the Purge & Trap/Gas Chromatograph system to detect biodegradation metabolites or changes in the hydrocarbon degradation rates (Figure 12).

Gas Chromatograph Analysis of Jet Fuel WSF

Sampling Regime

In all of the microcosm experiments, immediately following treatment with the water soluble fraction of the jet fuel, one replicate from each treatment group was collected for Purge and Trap/Gas Chromatograph (P&T/GC) analysis. Hydrocarbon component identification and quantification was performed with modifications according to USEPA Methods 601 (1982a) and 602 (1982b) (Westendorf, 1986). A clean, sterile 5 ml pipet was used to remove 5 ml from each treatment replicate to transfer and dispense the sample into a sterile, pre-cleaned 4 ml, 15 mm x 45 mm vial. Care was taken to minimize turbulence and potential release of volatiles. The vial was filled to prevent any headspace or air bubbles and sealed with a Teflon®-lined screw cap. All samples were stored in the dark at 4°C until the time of analysis.

In the MFC experiment samples were initially collected from the next replicate in each treatment group every two days, for a total of fourteen days. Purge and Trap/Gas Chromatograph analyses were completed within one week of sampling. In the re-treatment experiment, the MFC microcosm sampling frequencies were increased to every twelve hours with samples collected from all replicates at the same time and P&T/GC analyses conducted within two days of collection.

In the SAM experiment samples were collected every twelve hours and analyzed within twenty-four hours of collection on the P&T/GC system for both the initial and the re-treatment experiments. Sampling durations were also extended to one month, even though all initial volatiles had disappeared within two weeks of treatment.

Purge and Trap/Gas Chromatograph

A Tekmar® LSC 2000 Purge and Trap (P&T) concentrator system in tandem with a Hewlett-Packard® 5890A Gas Chromatograph (GC) with a Flame Ionization Detector (FID) was used for the analysis of all standards and microcosm samples (Appendices B-

C). Instrument blanks and deionized distilled water blanks were used to verify the P&T and GC columns cleanliness prior to analysis of samples.

On the day of analysis, samples were allowed to reach ambient temperature in the dark prior to analysis. A 5 ml gas tight Teflon® Luer lock Hamilton® syringe was used to remove a 3.5 ml sample from the 4 ml sample vial, using care to minimize any turbulence or incorporation of air into the sample, and injected into the 5 ml glass sparger on the P&T system (Appendix B). The sample was then purged with pre-purified nitrogen gas for eleven minutes to strip all the volatiles from the sample. The volatiles were collected on the Tenax/Silica Gel column trap and were then dry purged for four minutes to remove excess water. The trap was then heated very rapidly to 180°C and the flow of the nitrogen was reversed to rapidly desorb and carry the volatile hydrocarbons directly onto the Gas Chromatograph's SPB-5, 30 m x 0.53 mm ID, 1.5 μm film, fused silica capillary column. The GC column was programmed to hold at 35°C for two minutes, increase to 225°C at 12°C/min and hold at that temperature for five minutes. A Spectra-Physics 4290 Integrator recorded the FID signal output of the volatile hydrocarbons that had separated and eluted from the column by molecular weight and boiling point (Figures 7-12). A comparison was then made of the sample chromatograph peak retention times and area under the peak curve to *n*-paraffin and *aromatic chromatograph* reference standards that were prepared and analyzed under the same conditions for sample concentration determinations (Appendix C).

Statistical Analysis

Data were reported as area under the peak curve for the water soluble components, and were logarithmically transformed (base 10) to insure variance homogeneity and improve linear regression techniques. The logarithmic data points were regressed for each hydrocarbon component in each of the treatment groups. The slope of the regressed line for each of the hydrocarbons in each of the treatment groups was used as the rate of degradation of the component in time. The individual slope coefficients were compared in each of the other treatment groups and between both microcosms for significant differences in degradation rates using the Student's t test statistic (Tables 10-12, 16-19) (Zar, 1984).

Results

In both the Mixed Flask Culture (MFC) microcosm and the Standardized Aquatic Microcosm (SAM) experiments, three treatment groups were amended with serial concentrations of water solubilized fractions of hydrocarbons that were derived from the jet fuels JET-A and JP-8, respectively. The degradation rates of the hydrocarbon components were monitored using the Purge and Trap/Gas Chromatograph (P&T/GC). Each chromatograph peak represented a specific hydrocarbon component. The magnitude of each peak was a function of the concentration of the volatile hydrocarbon in the sample (Figures 7-12). The chromatographs enabled the comparison of the degradative rates for the individual hydrocarbon components between the treatment groups, within each microcosm experiment and between the treatment groups of the two microcosms.

To increase the linearity of the degradation slopes for individual hydrocarbons and to increase the homogeneity of their variances all chromatograph peak areas were transformed to logarithmic scale and regressed (Appendices D and E) (Zar, 1984). The slopes of their regressed concentrations through time were used to estimate rates of degradation. The greater the absolute value of the slope above zero, the faster the rate of degradation. In both the MFC and SAM experiments the slopes of selected hydrocarbons for each of the treatment groups were plotted on one graph (Figures 19-31). The mean area under the peak curve values and standard deviations for the re-treated microcosms were also included.

The slopes of the individual hydrocarbon components were compared, using the Student's t test to determine whether their degradation rates in the 15% WSF treatment group were significantly different from their degradation in the re-treated 0%+15% and the 15%+15% WSF groups (Tables 10-12) (Zar, 1984). In addition, the degradation rates of the hydrocarbon components in the 0%+15% and the 15%+15% WSF treatment groups were compared to each other to determine whether they were significantly different (Tables 10-12).

The specific hydrocarbon components that were the focus of this study were selected on the basis of their presence in both the JET-A and the JP-8 water soluble fractions. The hydrocarbon components selected were categorized by their chemical structure into three major classes of hydrocarbons the alkanes, the aromatics and the alkyl-substituted aromatics. The alkanes consisted of decane, dodecane, tridecane, and tetradecane (Figure 1). The aromatics consisted of benzene, toluene, and *meta*-, *ortho*-, and *para*-xylenes (Figure 2). The alkyl-substituted aromatics were ethylbenzene, propylbenzene, and butylbenzene (Figure 2). Cyclooctane was also included in the group

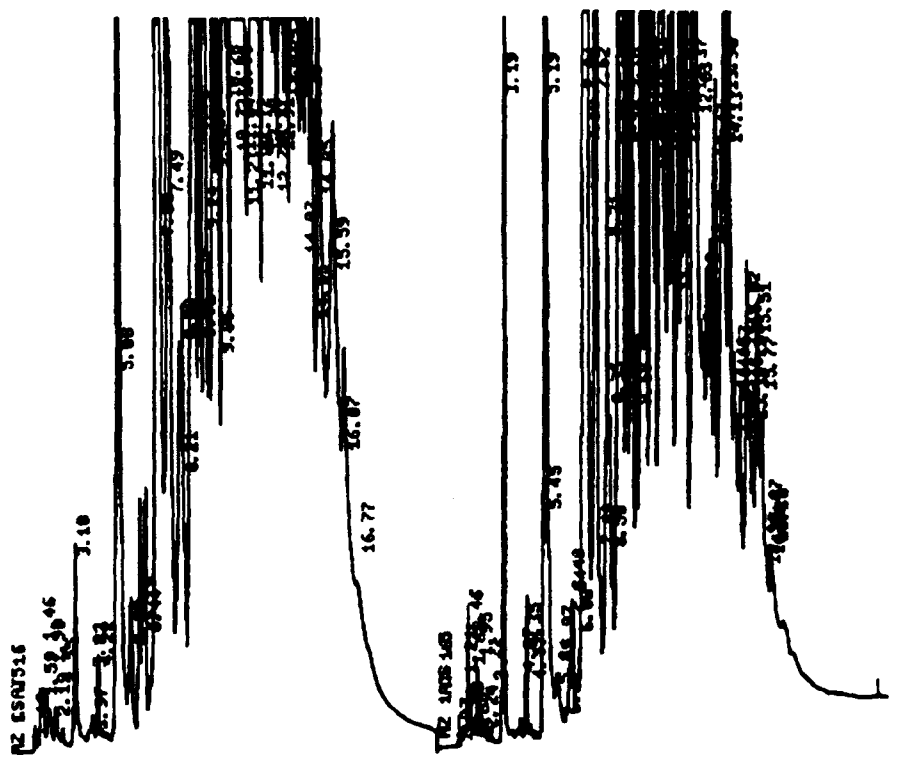


Figure 8. Gas Chromatographs of JP-8 15% WSF (left) and JET-A 15% WSF (right), immediately after treatment (0.0 hours).

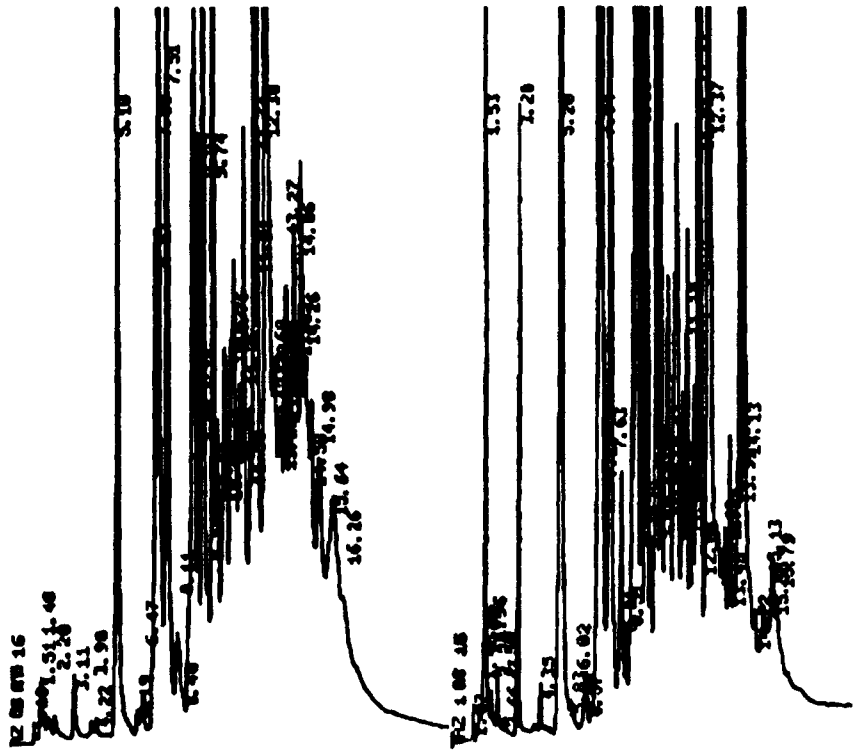


Figure 9. Gas chromatographs of JP-8 15% WSF (left) and JET-A 15% WSF (right), approximately 48 hours after treatment.

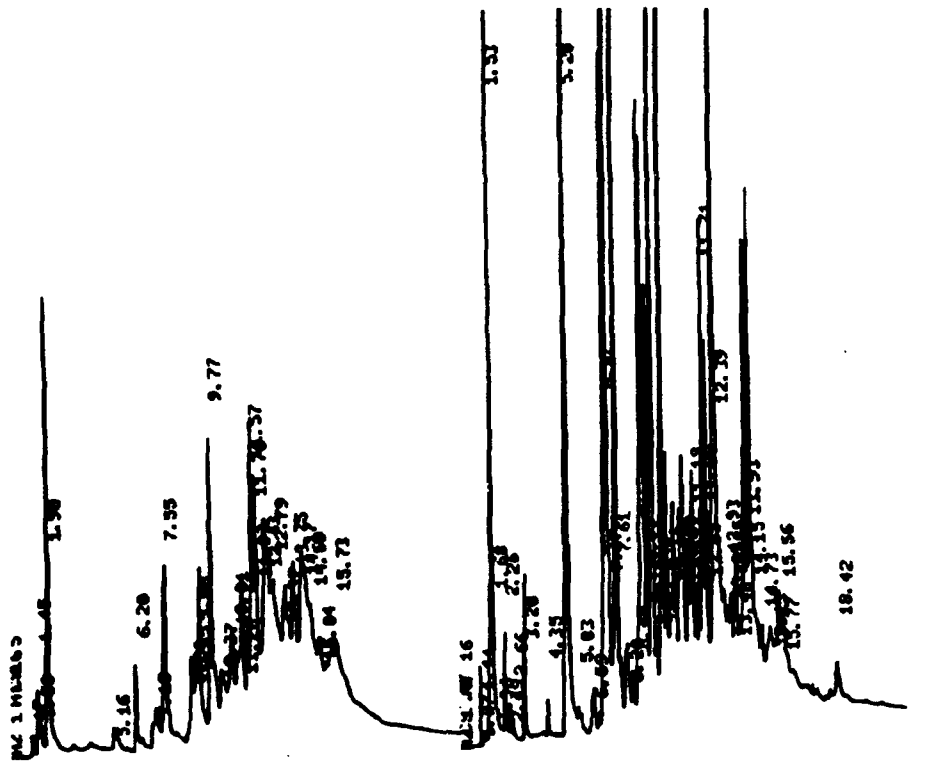


Figure 7. Gas Chromatographs of JP-8 15% WSF (left) and JET-A 15% WSF (right), approximately 96 hours after treatment.

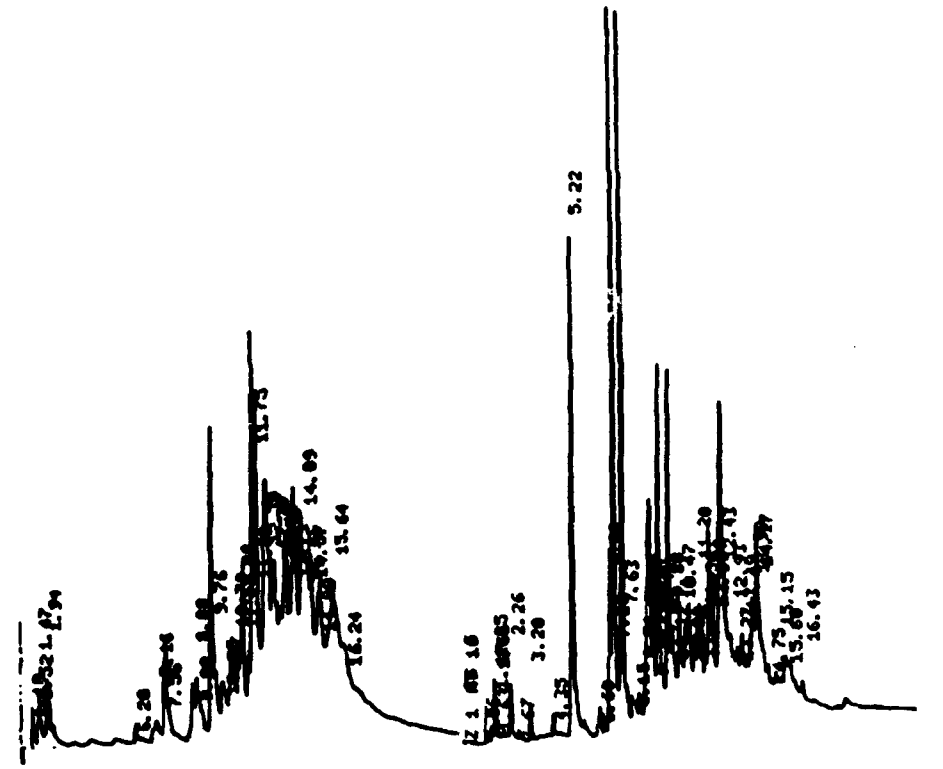


Figure 8. Gas Chromatographs of JP-8 15% WSF (left) and JET-A 15% WSF (right), approximately 144 hours after treatment.

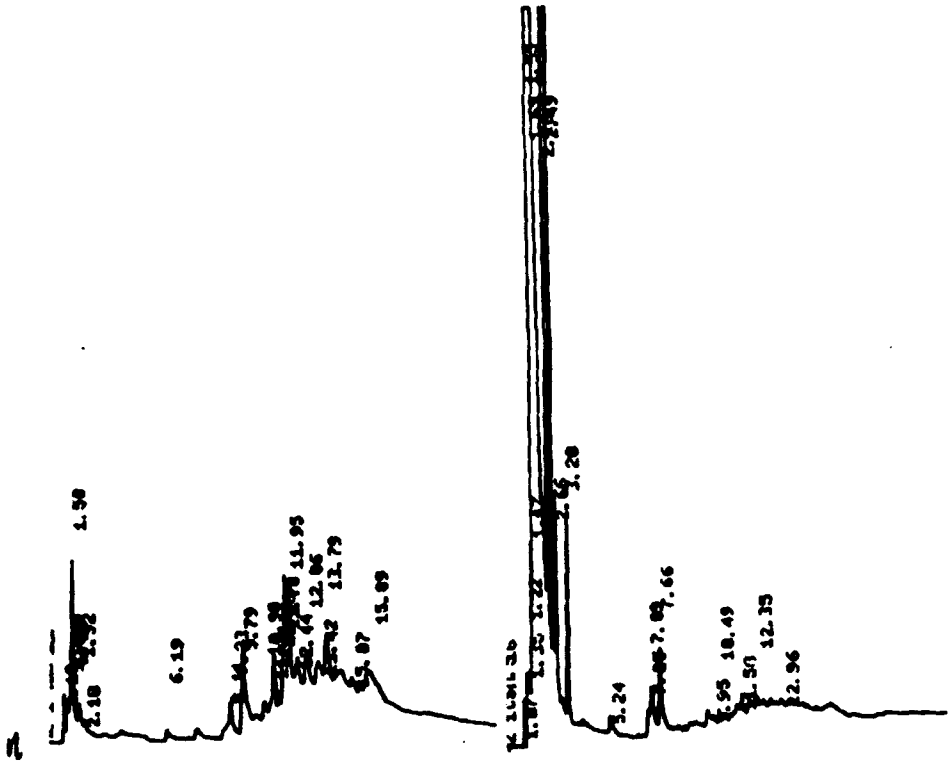


Figure 86. Gas Chromatographs of JP-8 15% WSF (left) and JET-A 15% WSF (right), approximately 208 hours after treatment.

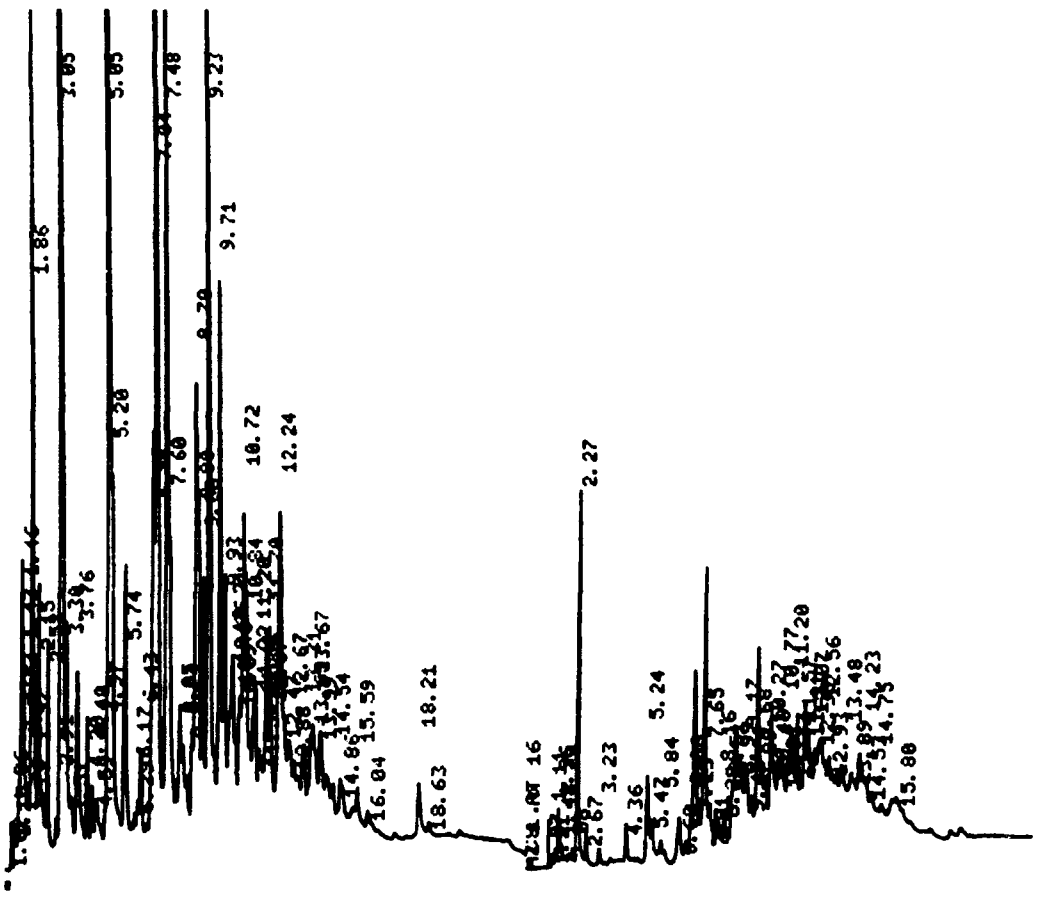


Figure 9. Gas Chromatograph of a sterilized MFC treated with JET-A 15% WSF (left), 190 hours after treatment compared to an unsterilized MFC treated with JET-A 15% WSF (right), 183 hours after treatment.

of alkyl-substituted aromatics.

All other hydrocarbons that were present or appeared during the course of the experiments were considered metabolic by-products or intermediates formed during the degradative process. These compounds were 2,4-dimethylpentane, 2-methylpropane, pentane, *cis*-2-pentene, *trans*-2-pentene, and propane. Hexane, butane, and 3-methylpentane were present as minor components in the jet fuel(s) and also as metabolites of hydrocarbon degradation. These hydrocarbons were not emphasized in the general analysis of hydrocarbon degradation processes (Figures 31-38).

Modifications were made in the experimental design prior to the initiation of the re-treatment experiment in the JET-A MFC microcosm, and were later implemented in the JP-8 SAM experiments. The modifications involved increasing the sampling frequency from every forty-eight hours to every twelve hours, decreasing the sample holding time prior to analysis from four days to a maximum of two days, and including more replicates in the analyses to improve statistical robustness. As a result, hydrocarbon degradation rates were monitored more precisely for longer periods of time and oscillations in the production of metabolites were revealed in greater detail during the course of the experiment. The presence of *cis*-2-pentene was originally not detected in the previous 1%, 5%, and 15% JET-A WSF treatments. However, in both the 0%+15% WSF and the 15%+15% WSF treatments the production of *cis*-2-pentene was detected to be a major metabolite of JET-A WSF degradation and occurred at approximately the same time intervals in both treatment groups (Figures 15a and 15b).

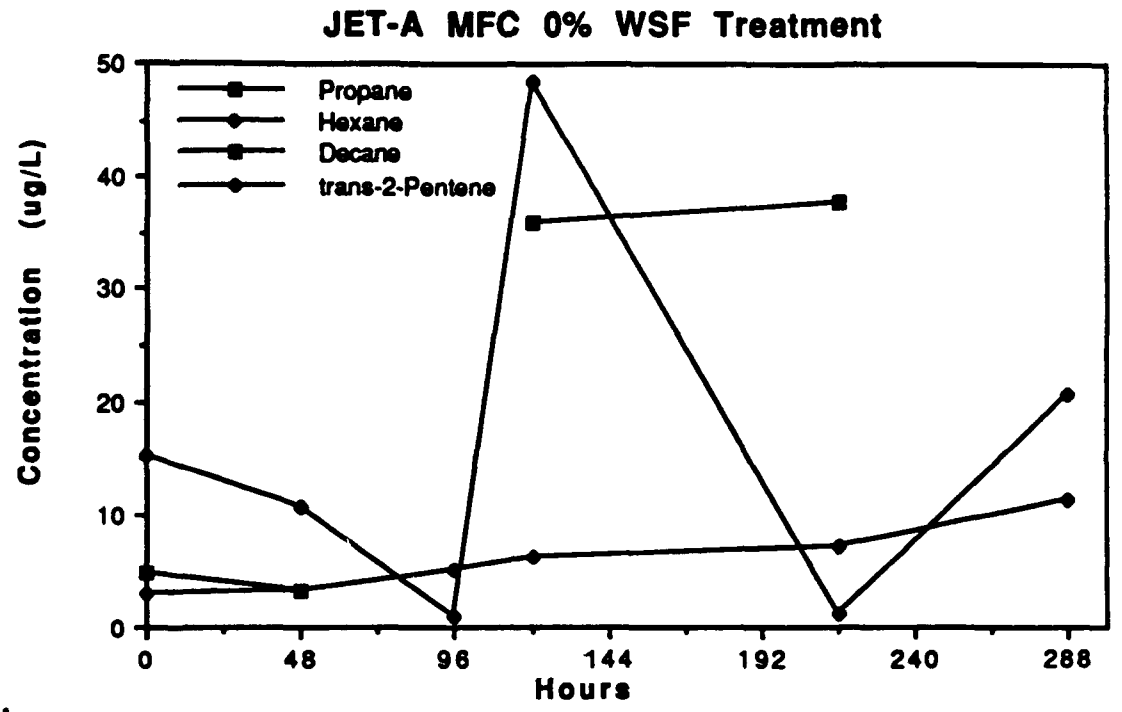
JET-A MFC and JP-8 SAM Results

In both the MFC and the SAM microcosm experiments there were several similar patterns in degradation rate responses and metabolite production. All hydrocarbon components diminished during the course of the experiments at relatively constant rates (Figures 19-30). Their degradation rates were linear functions with the duration of the degradation being dependent on the initial concentration of the water soluble fraction treatment. In the 1% WSF[1] treatment the degradation of all the hydrocarbon components was complete within four days (Figure 13b). In the 5% WSF treatments the degradation of the major hydrocarbon components was complete within five to ten days of treatment (Figures 14a and 17a). In the 15% WSF treatments the degradation of the major components was complete within twelve to fourteen days (Figures 14b and 17b).

---

[1]The results from the JP-8 SAM 1% WSF treatment microcosms were not utilized in this study, due to the loss of sample integrity from the prolonged holding period prior to analysis.

a.



b.

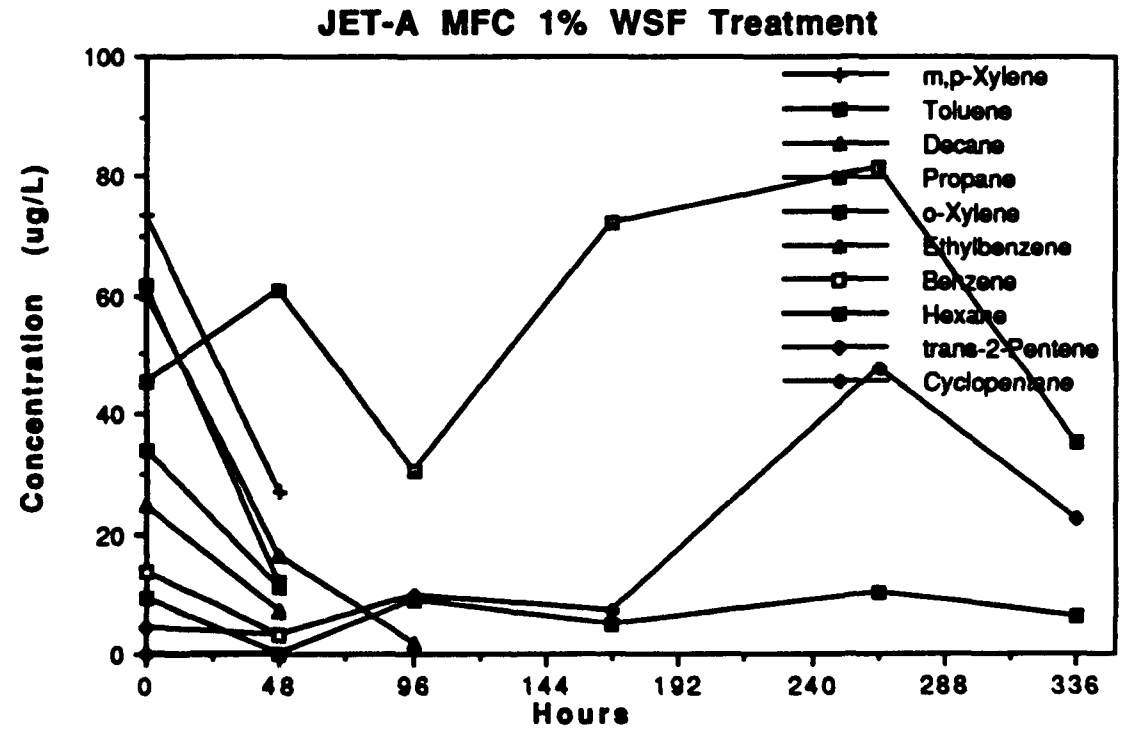
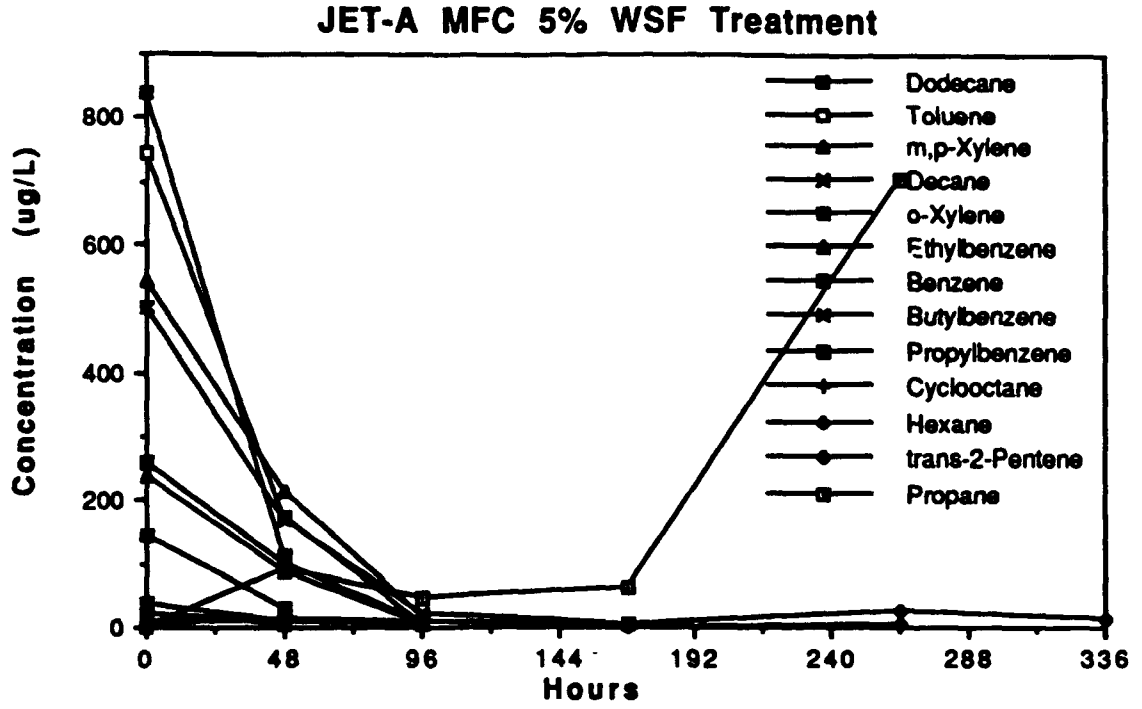


Figure 13. Hydrocarbon component concentrations (µg/L) for (a) the 0% WSF and (b) the 1% WSF JET-A MFC microcosm treatment groups.

a.



b.

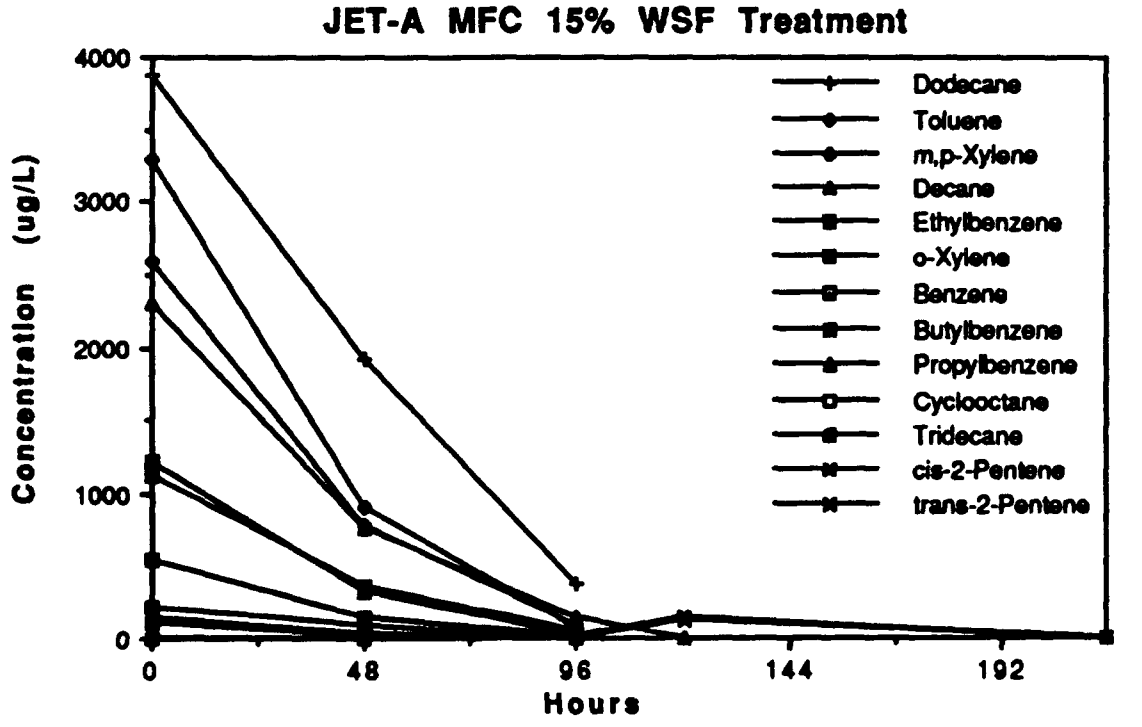
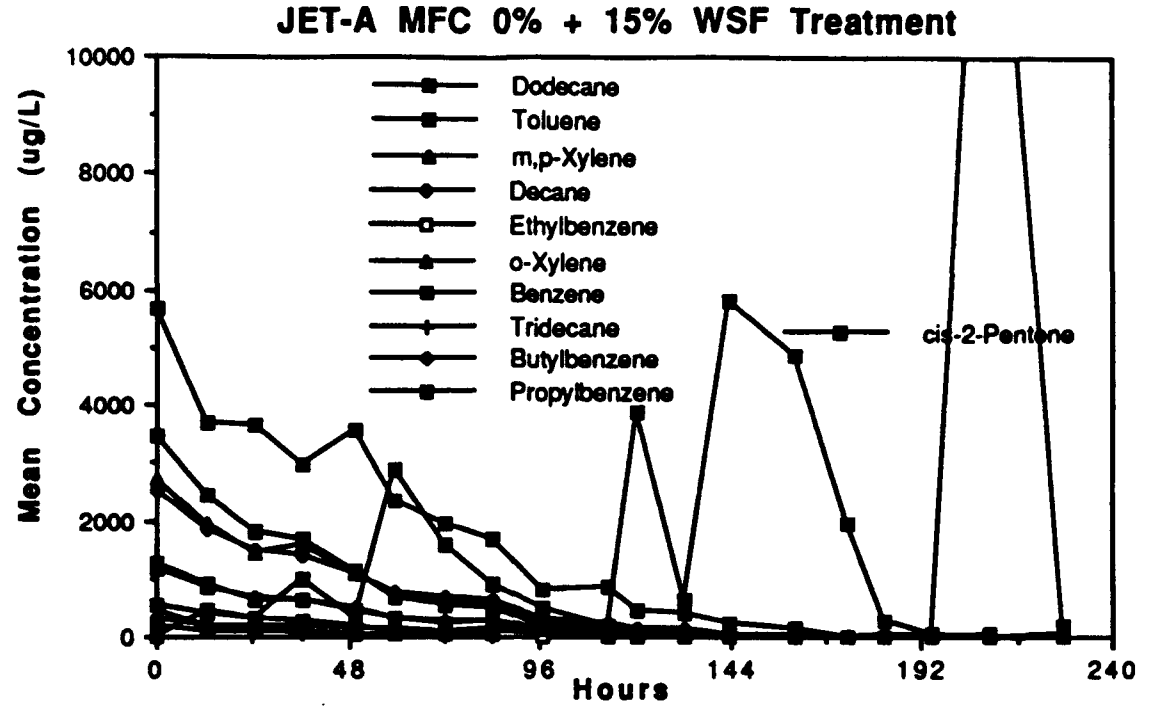


Figure 14. Hydrocarbon component concentrations (µg/L) for (a) the 5% WSF and (b) the 15% WSF JET-A MFC microcosm treatment groups.

a.



b.

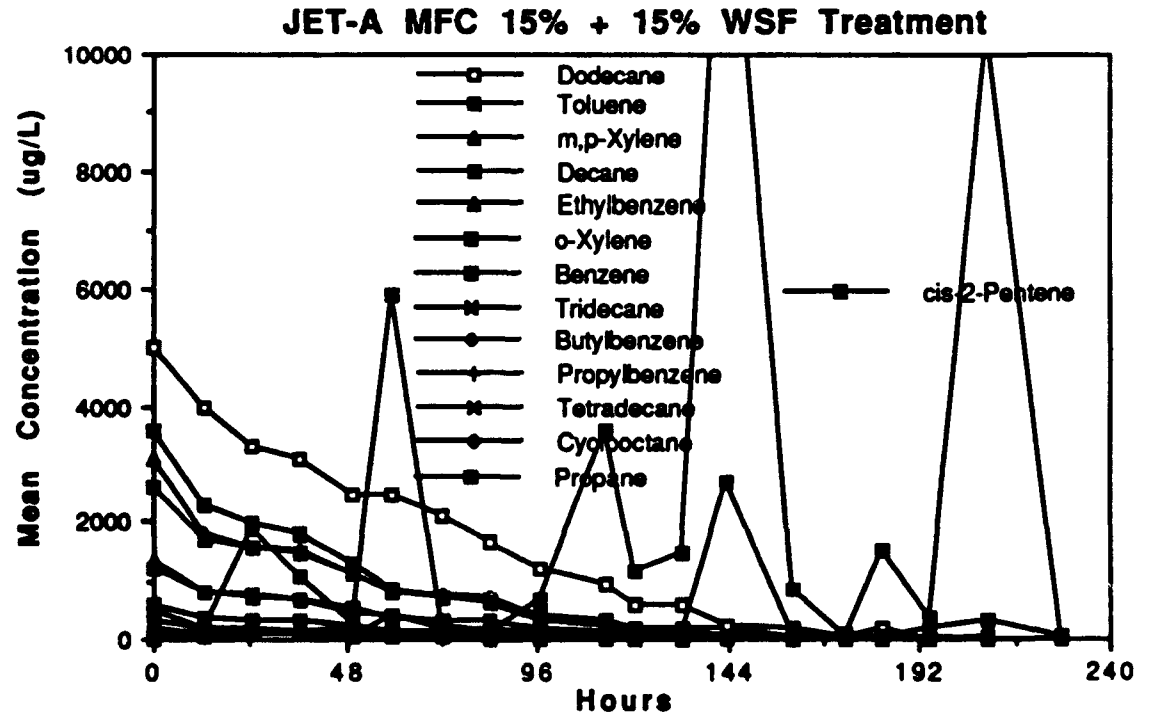


Figure 15. Mean hydrocarbon concentrations (µg/L) for (a) the 0% + 15 WSF and (b) the 15% + 15% WSF JET-A MFC microcosm treatment groups.

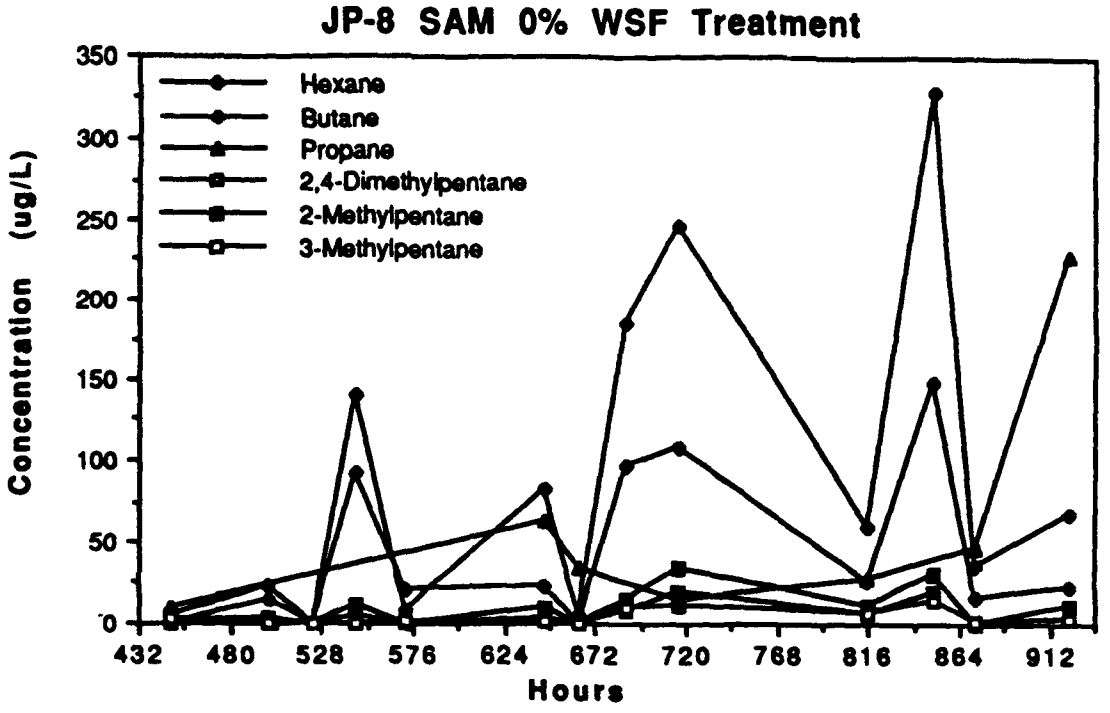
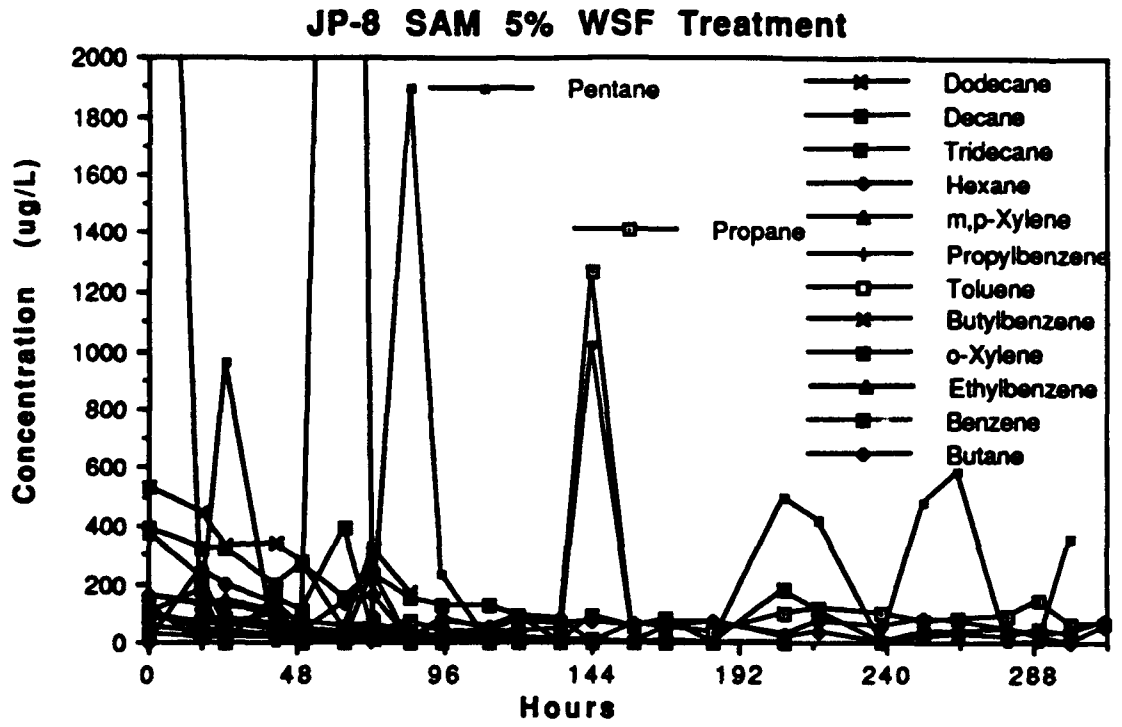


Figure 16. Hydrocarbon component concentrations (µg/L) for the 0% WSF JP-8 SAM treatment group.

a.



b.

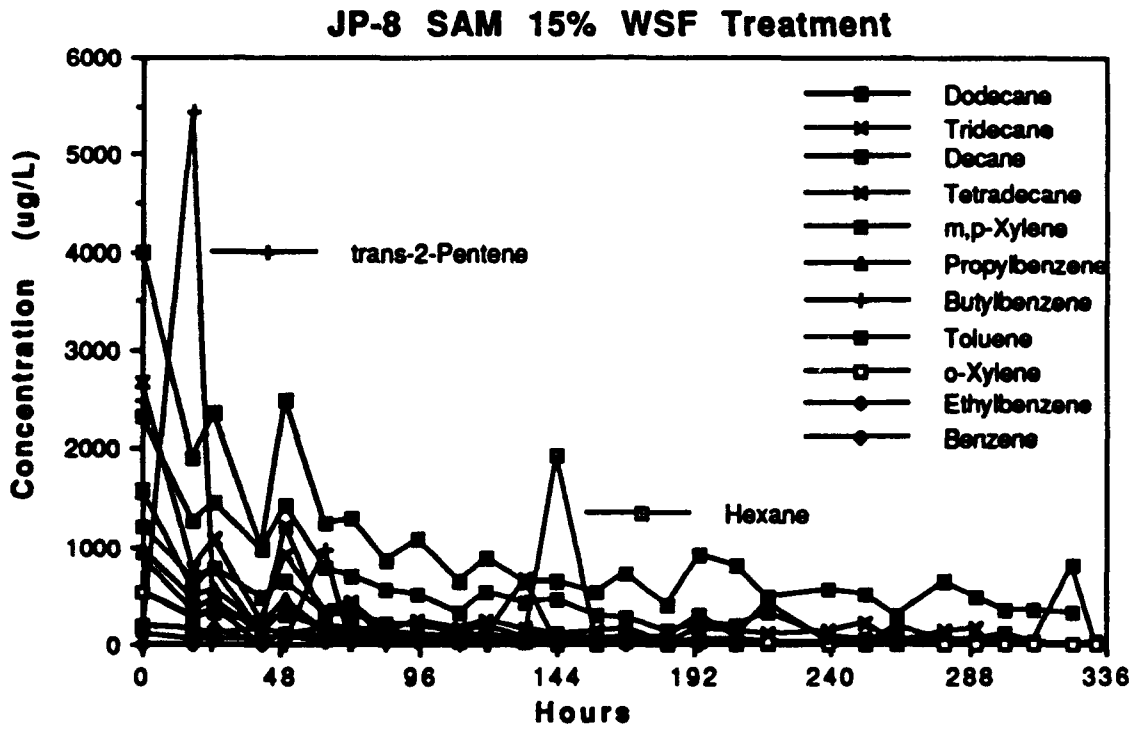
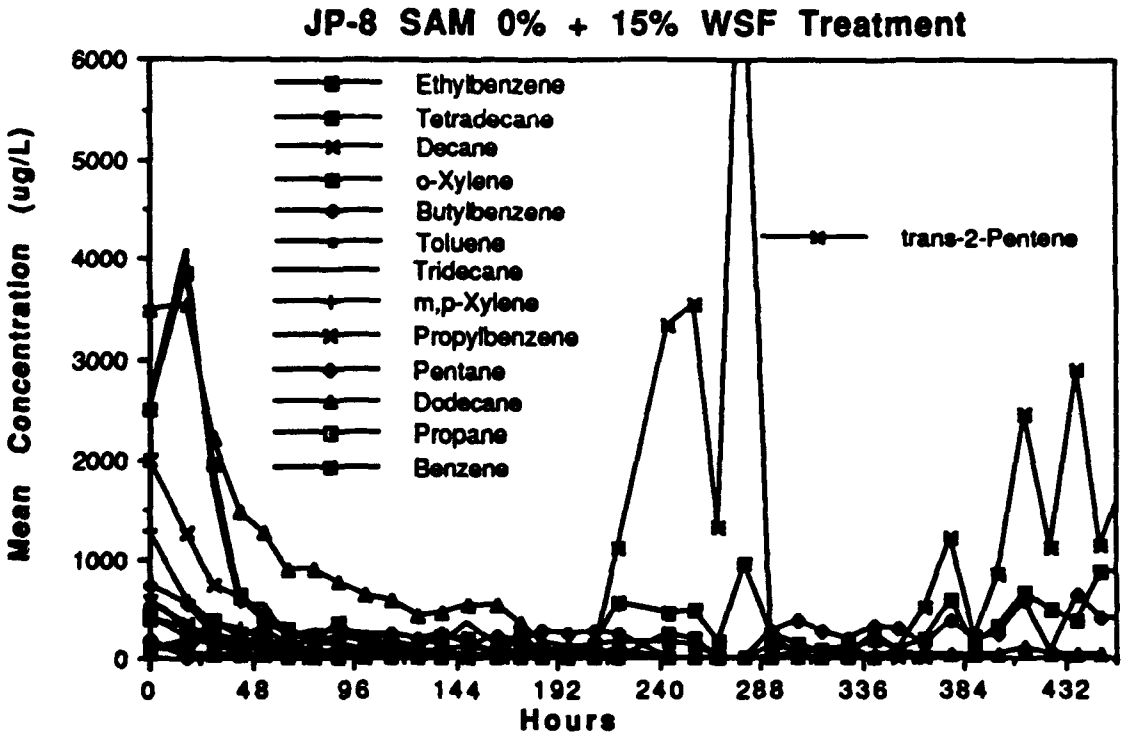


Figure 17. Hydrocarbon component concentrations (µg/L) for (a) the 5% WSF and (b) the 15% WSF JP-8 SAM microcosm treatment groups.

a.



b.

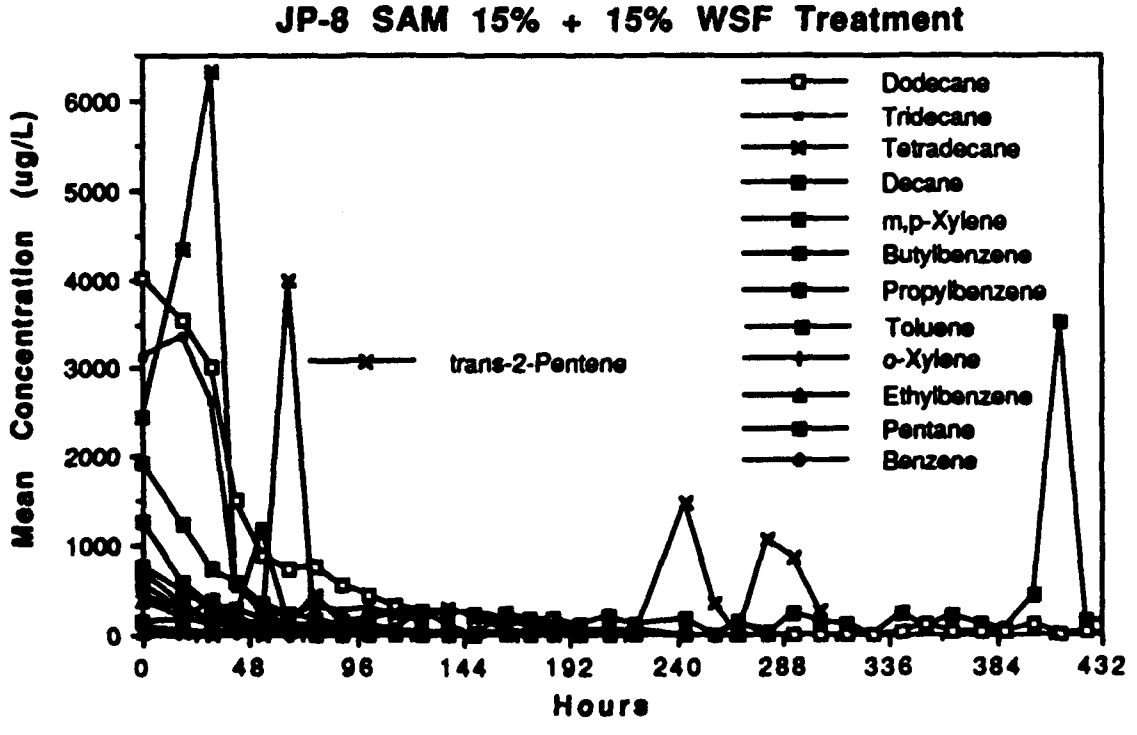


Figure 18. Mean hydrocarbon concentrations (µg/L) for (a) the 0% + 15% WSF and (b) the 15% + 15% WSF JP-8 SAM microcosm treatment groups.

Chromatographs of the respective 15%+15% WSF's of the two jet fuels in the MFC and the SAM displayed unique degradation profiles (Figures 7-12). At the initiation of the treatments the JP-8 water soluble fraction appeared to contain components in higher concentrations than in the JET-A WSF (Figures 7 and 8). A comparison of the chromatographs analyzed at set times during the course of their respective experiments showed that JET-A components were retained for longer periods of time (Figures 9 and 10). The components in highest concentration were identified to be the aromatics toluene, ethylbenzene, benzene and the xylenes (Table 2). The JP-8 components in highest concentration were identified to be the alkanes dodecane, tridecane, decane, and tetradecane (Table 3). The alkanes in the JP-8 appeared to be eliminated from the fuel mixture at a faster rate than the JET-A aromatics, however this relationship was an artifact of the sensitivity of the GC column. The alkanes were present in lower concentrations, but the GC column was more sensitive to them and resulted in chromatograms for JP-8 that were initially greater in magnitude. The water soluble fractions of the jet fuels in both microcosms were persistent for approximately the same period of time regardless of composition (Figure 11).

In the MFC and SAM reference microcosms the low molecular weight *n*-alkanes were present in low concentrations (Figures 13a and 16). At the initiation of the MFC experiment the alkanes propane, hexane, pentane, 2,4-dimethylpentane and *trans*-pentene were readily detected (Figure 13a). Conversely, in the SAM experiment none of these alkanes were detected until approximately 18 days into the experiment (Figure 16). The hydrocarbons were similar to those detected in the MFC reference microcosms and included propane, hexane, butane, 2,4-dimethylpentane, 2-methylpentane, and 3-methylpentane. In both microcosm experiments these same compounds were also detected at relatively low concentration levels in the other treatment groups. The production of these alkanes generally increasing in magnitude, as the treatment concentration increased (Figures 13-18). The production of alkanes has been attributed to the degradative oxidation of biogenic aromatic compounds by bacterial and algal communities (Nalewajko, 1977).

The degradation of the water soluble fractions of both jet fuels were primarily through microbially-mediated oxidative transformation and utilization mechanisms. The production of low molecular weight alkanes, branched alkanes, alkenes and mono-cyclic aromatic compounds during the course of the two experiments that were not present in the initial water soluble fractions, were attributed to microbial degradative activities. In the sterile MFC experiment all hydrocarbon component degradations were severely reduced for at least one week. The sterilized microcosms were allowed to be

exposed to airborne microorganisms and after approximately two weeks the hydrocarbon degradation rates increased and were degraded completely by the end of one month (Figure 12). The sterile microcosms also produced hydrocarbon components after two weeks that were similar to the metabolites detected in the unsterilized MFC and SAM microcosms.

The degradation rates of the hydrocarbon components in the 1%, 5% and 15% WSF treatment groups were more similar to each other, within the same microcosm experiment, than to the rates of degradation in the re-treated 0%+15% and the 15%+15% WSF treatment experiments. The initial rate of degradation of a specific hydrocarbon component at one treatment level was consistently the same or similar to its initial rate of degradation in the other two treatment groups. The degradation slopes for decane, dodecane, toluene, benzene, the xylenes, ethylbenzene, cyclooctane, and butylbenzene illustrate this relationship (Figures 19-20 and 23-29). As the concentration of the hydrocarbon decreased, the rate of degradation changed as well. The slope of the hydrocarbon component degradation in each of the treatment groups diverged to produce final regression equations that were not indicative of the initial period of similarity (Appendices D-E).

In the 0%+15% and the 15%+15% WSF treatments the initial slopes of the degradation curves began at the same concentration level on the y-intercept. In the JET-A WSF re-treated groups the degradation slopes of decane, benzene, *o*-xylene, butylbenzene and propylbenzene rapidly diverged to be significantly different from the 15% WSF treatment slopes (Tables 10-12). In the JP-8 re-treated groups the slopes of decane, dodecane, benzene, toluene, *m,p*-xylene, *o*-xylene, ethylbenzene and propylbenzene diverged and became significantly different from the 15% WSF treatments (Tables 10-12).

In both microcosms the degradation rates of the individual hydrocarbon components in the 0%+15% WSF treatments were not significantly different from their matching components in the 15%+15% WSF treatments during the first two to fours days after treatment. As degradation continued, the individual hydrocarbon degradation rates in the 0%+15% WSF also diverged from the 15%+15% WSF treatment rates to become significantly different (Figures 39-50). In the JET-A MFC benzene, toluene, and *o*-xylene became significantly different with the 0%+15% WSF treatment components degraded at faster rates. In the JP-8 SAM *dodecane*, the xylenes and ethylbenzene became significantly different, but the 15%+15% WSF treatment components were degraded faster (Table 16) (Tables 10-12).

In the JET-A MFC microcosm experiment the degradation slopes of an individual

hydrocarbon component in the 1%, 5%, and 15% WSF treatments were generally parallel with each other and the 15% WSF treatment components were generally degraded faster than in the 0%+15% and 15%+15% WSF groups (Figures 19a - 30a). In the JP-8 SAM experiment the degradation slopes of an individual hydrocarbon in the 1%, 5%, and 15% WSF treatments were also generally parallel with each other, but the 15% WSF treatment components were consistently degraded at slower rates than in the 0%+15% and 15%+15% WSF treatment groups (Figures 19b - 30b).

JET-A MFC Results

The class of hydrocarbons that were present as a group in the highest concentrations in all of the MFC microcosms were the aromatics (Table 2). Though dodecane was the single hydrocarbon component in the highest concentration, the aromatics toluene, *meta*-, *para*-, and *ortho*-xylenes, ethylbenzene, and benzene were the major constituents of the water soluble fractions. The other components that were present in slightly lower concentrations were the longer chain alkyl-substituted aromatics butylbenzene and propylbenzene including the cycloalkane cyclooctane. The longer carbon chain *n*-alkane compounds decane, tridecane and tetradecane were present in the least amounts (Table 2).

The hydrocarbon components were separated into their respective chemical classifications and ranked by concentration levels, with the component in the highest concentration ranked first. In each of the treatment groups the same or similar hierarchical rankings were maintained except in the 1% WSF treatment (Tables 4a, 6a, and 8a). The alkane rank order was dodecane, decane, tridecane, and tetradecane. The 1% WSF treatment group was the exception with decane being the only alkane detected (Table 3a). The aromatics were toluene, *m,p*-xylene, *o*-xylene, and benzene with the exception being in the 1% WSF treatment where the order was *m,p*-xylene, toluene, *o*-xylene, and benzene (Table 6a). In the alkyl-substituted aromatics the order was ethylbenzene, butylbenzene, propylbenzene, and cyclooctane. In the 1% WSF treatment ethylbenzene and cyclooctane were the only components detected (Table 8a). Above the 1% WSF concentration level the hydrocarbon components were solubilized to the same extent and retained the same hierarchy in their concentrations. The re-treated 0%+15% and 15%+15% WSF treatments maintained the same hierarchical rankings in the alkane, aromatic, and alkyl-substituted aromatic concentrations as displayed in the single treatment groups (Tables 4a, 6a, and 8a).

The concentrations of the individual hydrocarbon components were higher in each successive treatment group as the percentage of the WSF treatment increased. However,

Table 2. Initial concentrations (µg/L) of the individual hydrocarbon components in the JET-A water soluble fraction treatment groups.

	JET-A MFC Treatment Group Concentrations (µg/L)					
Hydrocarbon	0% WSF	1% WSF	5% WSF	15% WSF	0%+15% WSF	15%+15% WSF
Dodecane			838.0	3874.9	5690.3	5043.3
Toluene		61.5	741.8	3294.1	3459.6	3599.0
m,p-Xylene		73.4	541.3	2590.2	2682.9	3103.8
Decane	4.8	60.4	502.3	2298.6	2525.0	2599.9
Ethylbenzene	15.2	24.8	236.8	1212.9	1269.2	1356.4
o-Xylene		33.9	257.0	1121.0	1170.2	1198.8
Benzene		13.6	145.7	539.8	556.7	572.2
Butylbenzene			35.5	209.1	305.5	327.7
Propylbenzene			21.6	140.0	158.5	170.6
Cyclooctane			21.6	114.8	120.6	121.8
Tridecane				102.2	450.0	498.9
Propane		45.3	88.5	47.0	159.9	111.4
Tetradecane				35.0	87.1	132.6
trans-2-Pentene	3.0	4.4	7.7	3.6	530.6	26.4
cis-2-Pentene					12.2	5.8
Octane				3.9	10.4	18.6
3-Methylpentane			1.9	7.2	8.5	8.4
Hexane		9.3	5.1	7.3	6.4	9.0
2-Methylpropane					2.6	3.0
2,4-Dimethylpentane				1.7	2.1	3.1
Pentane				47.8		

Table 4. Alkane hydrocarbon components in rank order of concentration for each treatment group in (a) the JET-A MFC and (b) the JP-8 SAM microcosms.

a. JET-A MFC Alkanes

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1	Decane	Dodecane	Dodecane	Dodecane	Dodecane
2		Decane	Decane	Decane	Decane
3			Tridecane	Tridecane	Tridecane
4					Tetradecane

b. JP-8 SAM Alkanes

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1		Dodecane	Dodecane	Dodecane	Dodecane
2		Decane	Tridecane	Tridecane	Tridecane
3		Tridecane	Decane	Tetradecane	Tetradecane
4		Tetradecane	Tetradecane	Decane	Decane

Table 5. Alkane hydrocarbon components in rank order of degradation for each treatment group in (a) the JET-A MFC and (b) the JP-8 SAM microcosms.

a. JET-A MFC Alkanes

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1	Decane	Dodecane	Decane	Tridecane	Dodecane
2		Decane	Tridecane	Dodecane	Tetradecane
3			Dodecane	Decane	Tridecane
4			Tetradecane		Decane

b. JP-8 SAM Alkanes

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1		Tetradecane	Tetradecane	Tetradecane	Tetradecane
2		Tridecane	Tridecane	Tridecane	Tridecane
3		Dodecane	Decane	Decane	Decane
4		Decane	Dodecane	Dodecane	Dodecane

Table 6. Aromatic hydrocarbon components in rank order of concentration for each treatment group in (a) the JET-A MFC and (b) the JP-8 SAM microcosms.

a. JET-A MFC Aromatics

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1	m,p-Xylene	Toluene	Toluene	Toluene	Toluene
2	Toluene	m,p-Xylene	m,p-Xylene	m,p-Xylene	m,p-Xylene
3	o-Xylene	o-Xylene	o-Xylene	o-Xylene	o-Xylene
4	Benzene	Benzene	Benzene	Benzene	Benzene

b. JP-8 SAM Aromatics

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1		m,p-Xylene	m,p-Xylene	m,p-Xylene	m,p-Xylene
2		Toluene	Toluene	Toluene	Toluene
3		o-Xylene	o-Xylene	o-Xylene	o-Xylene
4		Benzene	Benzene	Benzene	Benzene

Table 7. Aromatic hydrocarbon components in rank order of degradation for each treatment group in (a) the JET-A MFC and (b) the JP-8 SAM microcosms.

a. JET-A MFC Aromatics

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1	Toluene	Toluene	Toluene	Toluene	Toluene
2	Benzene	m,p-Xylene	Benzene	Benzene	m,p-Xylene
3	o-Xylene	o-Xylene	m,p-Xylene	m,p-Xylene	Benzene
4	m,p-Xylene	Benzene	o-Xylene	o-Xylene	o-Xylene

b. JP-8 SAM Aromatics

Rank Order	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1		Toluene	Toluene	Toluene	Toluene
2		m,p-Xylene	m,p-Xylene	m,p-Xylene	m,p-Xylene
3		o-Xylene	Benzene	Benzene	Benzene
4		Benzene	o-Xylene	o-Xylene	o-Xylene

Table 8. Alkyl-substituted aromatic hydrocarbon components (and Cyclooctane) in rank order of concentration for each treatment group in (a) the JET-A MFC and (b) the JP-8 SAM microcosms.

a. JET-A MFC Alkyl-Substituted Aromatics

Rank	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1	Ethylbenzene	Ethylbenzene	Ethylbenzene	Ethylbenzene	Ethylbenzene
2	Cyclooctane	Butylbenzene	Butylbenzene	Butylbenzene	Butylbenzene
3		Propylbenzene	Propylbenzene	Propylbenzene	Propylbenzene
4		Cyclooctane	Cyclooctane	Cyclooctane	Cyclooctane

b. JP-8 SAM Alkyl-Substituted Aromatics

Rank	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1		Propylbenzene	Propylbenzene	Butylbenzene	Butylbenzene
2		Butylbenzene	Butylbenzene	Propylbenzene	Propylbenzene
3		Ethylbenzene	Ethylbenzene	Ethylbenzene	Ethylbenzene
4		Cyclooctane	Cyclooctane	Cyclooctane	Cyclooctane

Table 9. Alkyl-substituted aromatic hydrocarbon components (and Cyclooctane) in rank order of degradation for each treatment group in (a) the JET-A MFC and (b) the JP-8 SAM microcosms.

a. JET-A MFC Alkyl-Substituted Aromatics

Rank	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1	Ethylbenzene	Ethylbenzene	Ethylbenzene	Ethylbenzene	Ethylbenzene
2		Propylbenzene	Cyclooctane	Cyclooctane	Cyclooctane
3		Butylbenzene	Butylbenzene	Propylbenzene	Propylbenzene
4		Cyclooctane	Propylbenzene	Butylbenzene	Butylbenzene

b. JP-8 SAM Alkyl-Substituted Aromatics

Rank	1% WSF	5% WSF	15% WSF	Mean 0% + 15% WSF	Mean 15% + 15% WSF
1	Ethylbenzene	Propylbenzene	Propylbenzene	Propylbenzene	Propylbenzene
2		Cyclooctane	Butylbenzene	Cyclooctane	Cyclooctane
3		Ethylbenzene	Cyclooctane	Ethylbenzene	Ethylbenzene
4		Butylbenzene	Ethylbenzene	Butylbenzene	Butylbenzene

the increase in their concentration levels were not consistent with the concentrations of the WSF amendments (Table 2). The concentrations of the individual components increased several orders in magnitude above the expected concentration levels as the percentage of the WSF added to the microcosms was increased. In the 5% WSF the concentration of xylenes were seven times higher and toluene twelve times higher than expected. In the 15% WSF the concentrations of the individual hydrocarbons were also elevated above the expected levels. However, the degree of difference between the expected and the measured concentrations was much less with most components being only four times greater. Propylbenzene and butylbenzene concentration levels were the exceptions by being six times higher. In the re-treated microcosms the concentrations of the hydrocarbon components in the 0%+15% were generally higher than in the initial 15% WSF group while the 15%+15% WSF treatments were consistently greater than the component concentrations in either the 15% WSF or the 0%+15% WSF treatment group (Table 2).

The individual hydrocarbons in each treatment group were also categorized into their chemical classes and ranked in order of their decreasing rates of degradation (Tables 5a, 7a and 9a). A comparison of the ranked order of hydrocarbon concentrations to their ranked rates of degradation revealed that the two rankings were not the same for most of the treatment groups (Tables 4a - 9a). The *n*-alkanes in the 1% and 5% WSF were the exceptions. The concentration of decane was highest in the 1% WSF and it was degraded at the fastest rate (Tables 4a and 5a). In the 5% WSF dodecane was present in the highest concentration followed by decane and the order of degradation rates was dodecane first and decane second.

At the 15% WSF concentration level that included the 0%+15% and the 15%+15% WSF treatments the rank order of degradation was more variable. The trend seemed to indicate that the higher molecular weight alkanes which were in the lowest concentrations were degraded at the fastest rates, while the lower molecular weight alkanes degraded at slower rates (Tables 5a). These results are consistent with alkane utilization by microorganisms. The higher molecular weight alkanes in the WSF were preferential carbon substrates for the bacteria and were rapidly utilized and degraded. The lower molecular weight alkanes were present as metabolites and may have been less energetically useful or inhibitory to the microorganisms. These alkanes that were not utilized or were utilized at much slower rates accumulated in the microcosm medium.

A comparison of the ranked aromatic concentrations and degradation rates display slightly different dynamics than the alkanes. Toluene was highest in concentration in the 5%, 15%, 0%+15%, and the 15%+15% WSF treatments and was degraded at the

fastest rate in all treatment groups (Tables 6a and 7a). The rank order of hydrocarbon concentrations remained the same from the 5% WSF through the 15%+15% WSF treatments, while the rank order of hydrocarbon degradation rates did not remain consistent. The 15% and the 0%+15% WSF treatments had the same rank order of degradation, but the 15%+15% WSF treatment did not. The only consistent results were for toluene which was degraded at the fastest rate and for *o*-xylene which was degraded at the slowest rate (Tables 6a and 7a).

The alkyl-substituted aromatics were similar to the aromatics in that ethylbenzene which was present in the highest concentration was degraded at the fastest rate (Tables 8a and 9a). Cyclooctane which was in the lowest concentration in all treatment groups was degraded second to ethylbenzene in the 15%, 0%+15% and 15%+15% WSF treatments. As in the alkanes and the aromatic rankings, the ranked order of degradation rates were not consistent between any of the treatment groups. Unlike the aromatics, no individual compound was consistently degraded at the slowest rate.

Hydrocarbon Treatment Comparisons

In the alkanes the degradation of decane was the most rapid in the 15% WSF treatment, decreased in the 1% WSF, and was the slowest in the 5% WSF treatment (Appendix D)(Figure 19a). In both the 0%+15% and the 15%+15% WSF treatment groups decane degradation rates were significantly different from the 15% WSF treatment and were slower than in any of the other treatment group rates (Table 10). However, the rates of decane degradation in the 0%+15% WSF treatment compared to the 15%+15% WSF treatment were not significantly different.

Dodecane was degraded at the fastest rate in the 0%+15% WSF, followed by the 5% WSF, and 15%+15% WSF treatments, with the slowest rate in the 15% WSF treatments (Appendix D) (Figure 20a). There were no significant differences between the rates of degradation in the 15% WSF compared to the two re-treated groups or between the two re-treated groups (Table 10).

Tridecane and tetradecane were not detected in the lower treatment groups, but were present in the 15% WSF treatments and tetradecane was degraded faster than tridecane (Figures 21a and 22a). There were no significant differences between the rates of degradation in the 15% WSF compared to the two re-treated groups or between the two re-treated groups for both components (Table 10).

In the aromatic group of hydrocarbon components toluene was consistently degraded at the fastest rate in all of the treatment groups. The rate of degradation in the

Table 10. Student's t test for significant differences ($*t_{0.05\ (2),\ 28}$) between log regressed degradation rate coefficients for each alkane hydrocarbon component in the 15%, 0%+15% and the 15%+15% WSF treatments.

	JET-A MFC (WSF)		JP-8 SAM (WSF)	
Alkane (WSF)	0%+15%	15%+15%	0%+15%	15%+15%
Decane [15%]	*	*	*	*
Decane [0% + 15%]				
Dodecane [15%]			*	*
Dodecane [0% + 15%]				*
Tridecane [15%]				
Tridecane [0% + 15%]				
Tetradecane [15%]				
Tetradecane [0% + 15%]				

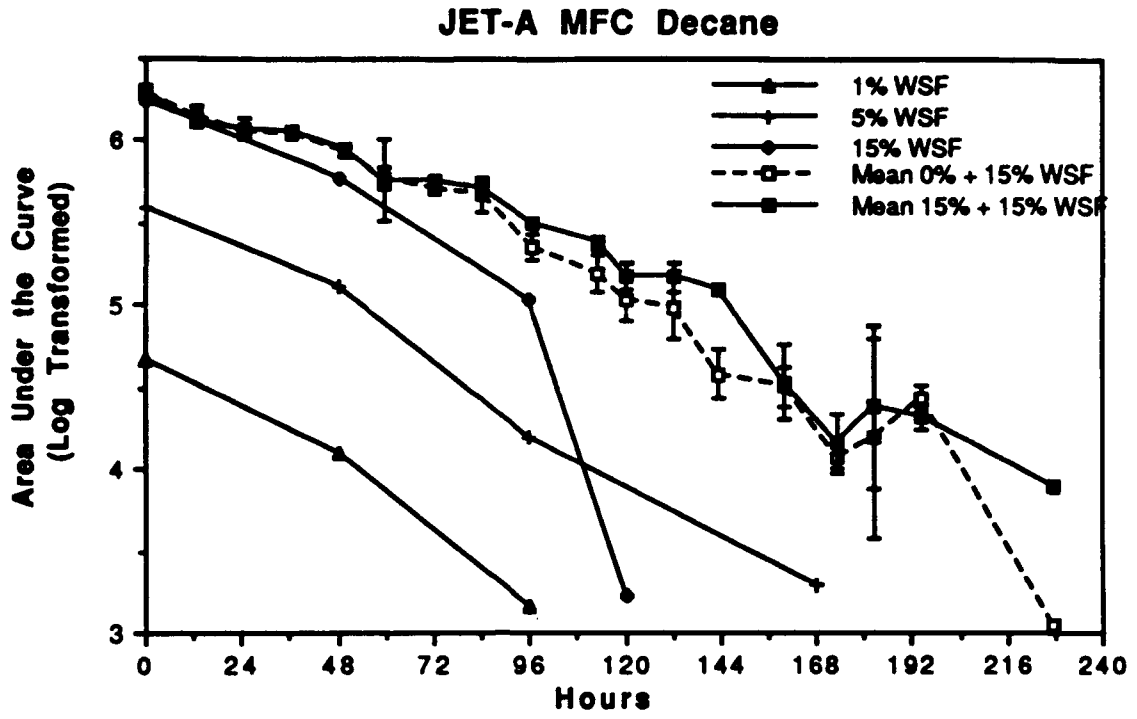
Table 11. Student's t test for significant differences ($*t_{0.05\ (2),\ 28}$) between log regressed degradation rate coefficients for each aromatic hydrocarbon component in the 15%, 0%+15% and the 15%+15% WSF treatments.

	JET-A MFC (WSF)		JP-8 SAM (WSF)	
Aromatic (WSF)	0%+15%	15%+15%	0%+15%	15%+15%
Benzene [15%]	*	*	*	*
Benzene [0% + 15%]		*		
Toluene [15%]			*	*
Toluene [0% + 15%]		*		
m,p-xylene [15%]			*	*
m,p-xylene [0% + 15%]				*
o-xylene [15%]		*	*	*
o-xylene [0% + 15%]		*		*

Table 12. Student's t test for significant differences ($*t_{0.05\ (2),\ 28}$) between log regressed degradation rate coefficients for each alkyl-aromatic hydrocarbon component in the 15%, 0%+15% and the 15%+15% WSF treatments.

	JET-A MFC (WSF)		JP-8 SAM (WSF)	
Alkyl-Aromatic (WSF)	0%+15%	15%+15%	0%+15%	15%+15%
Butylbenzene [15%]	*	*		
Butylbenzene [0% + 15%]				
Cyclooctane [15%]				
Cyclooctane [0% + 15%]				
Ethylbenzene [15%]			*	*
Ethylbenzene [0% + 15%]				*
Propylbenzene [15%]	*			*
Propylbenzene [0% + 15%]				

a.



b.

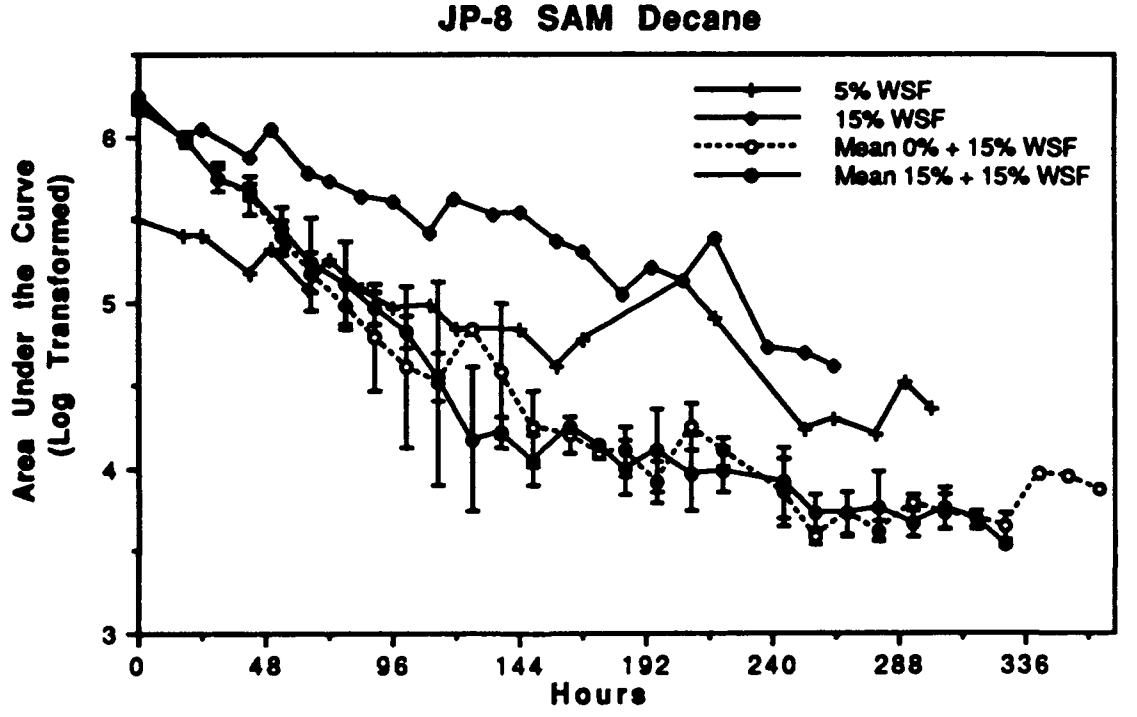
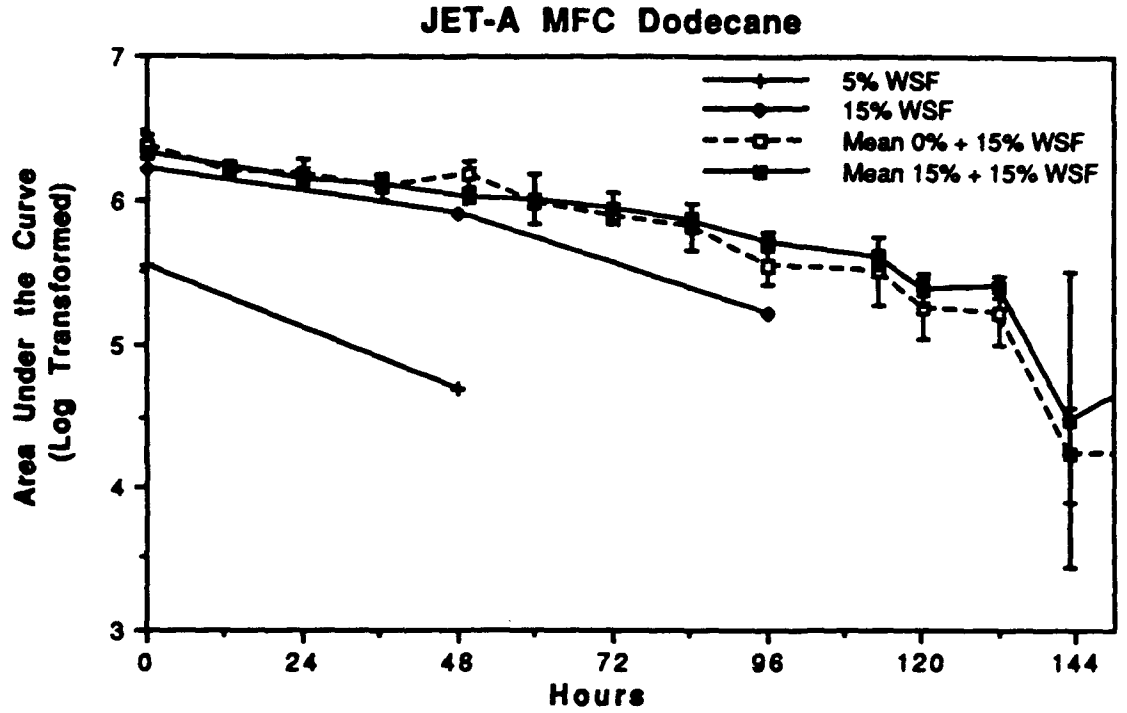


Figure 19. Log transformed degradation slopes for the *n*-alkane decane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

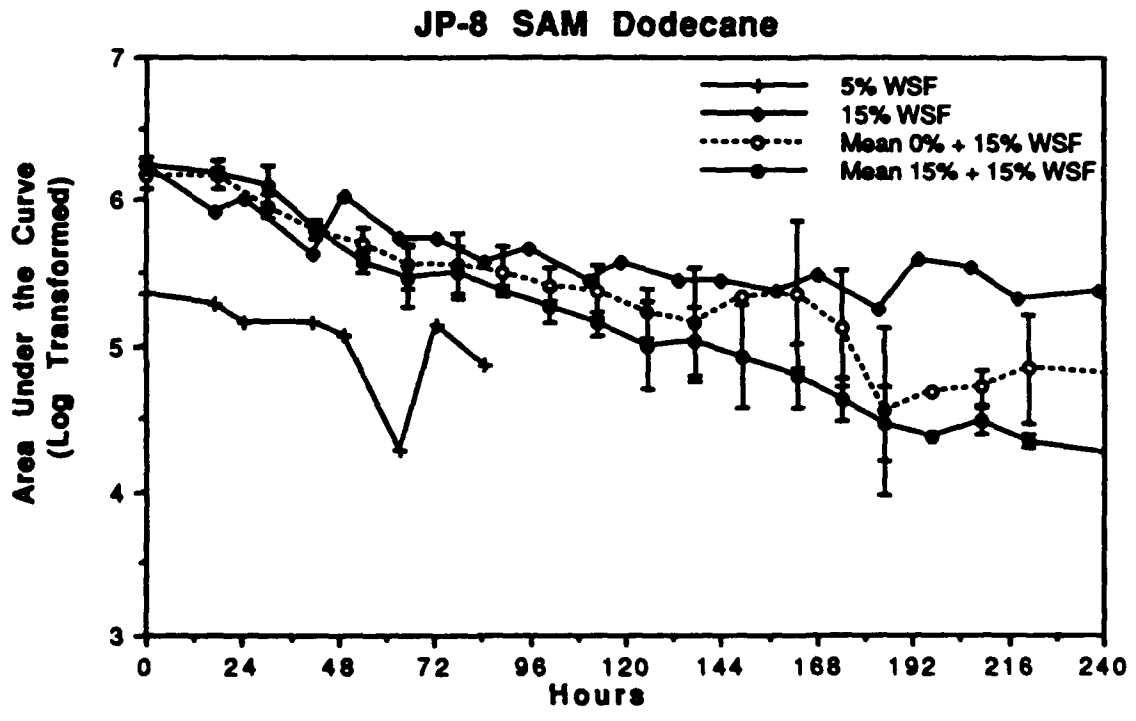
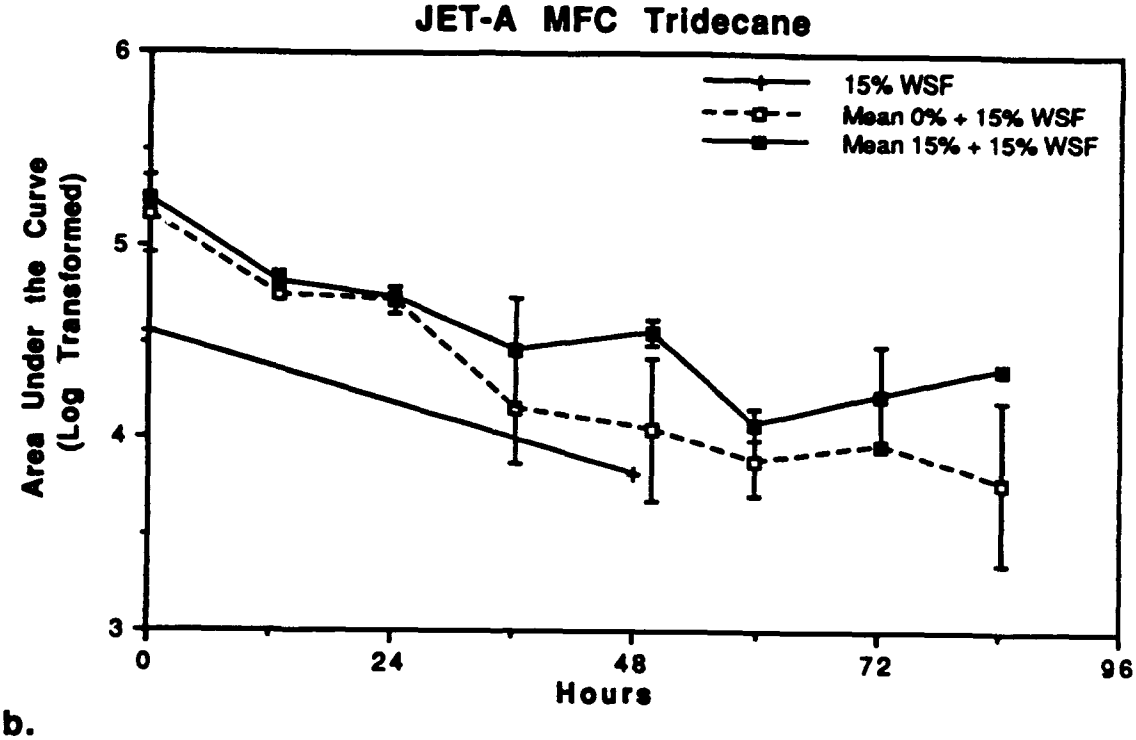


Figure 20. Log transformed degradation slopes for the *n*-alkane dodecane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

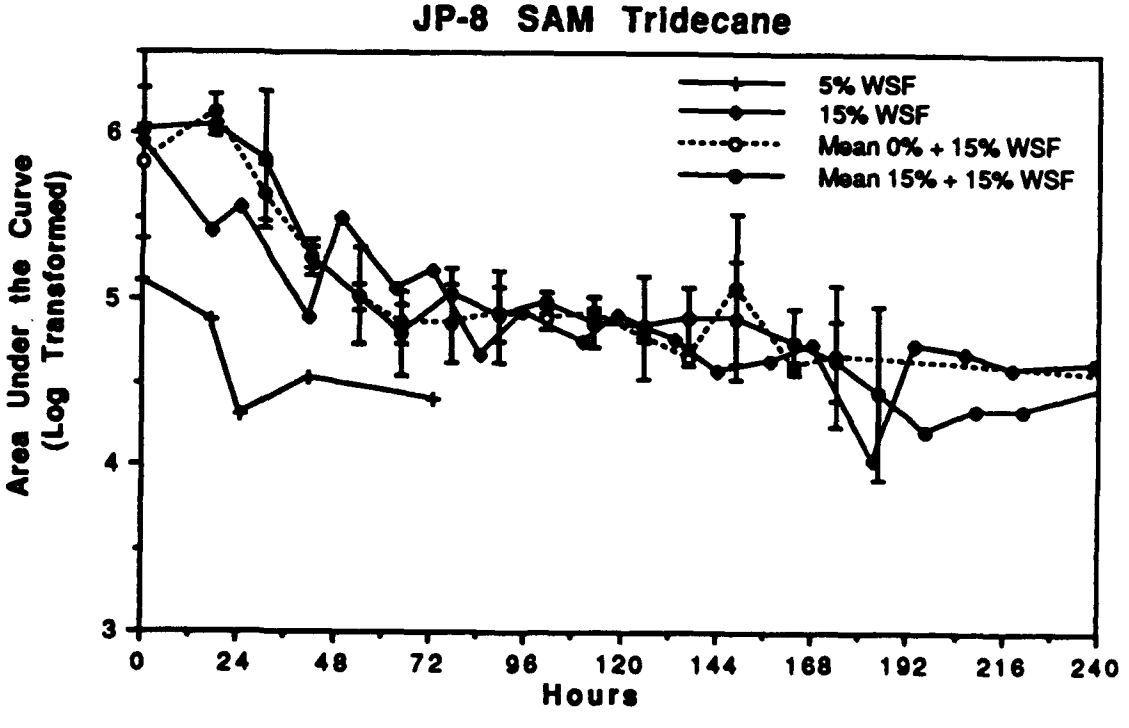
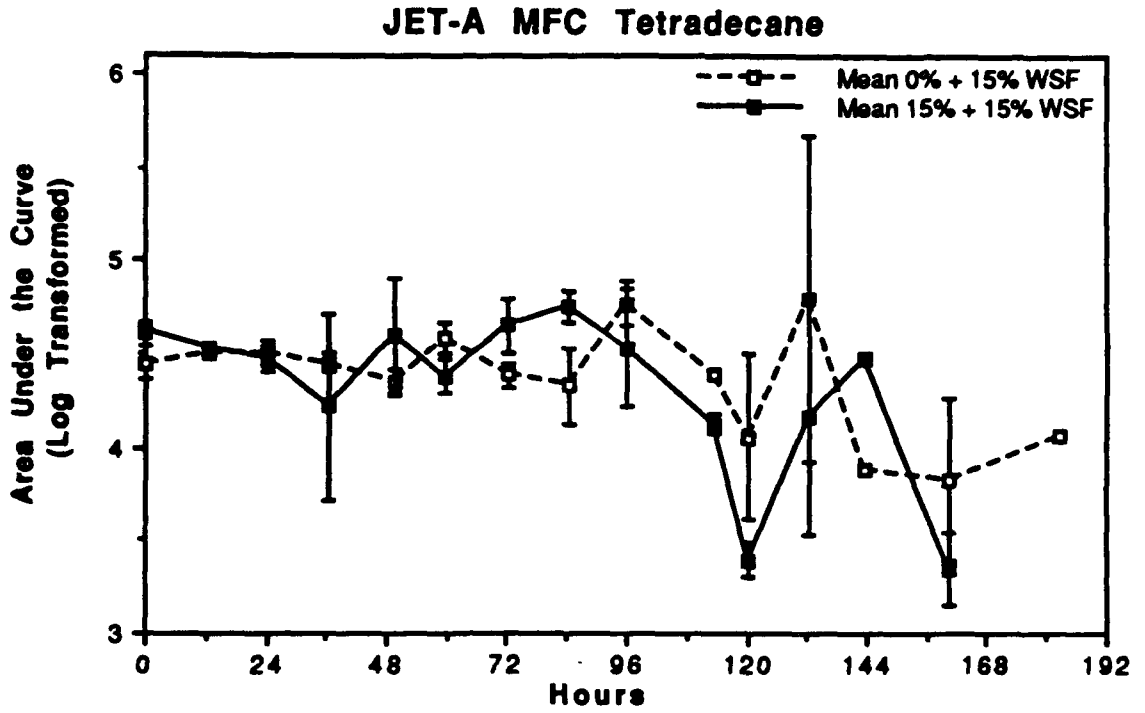


Figure 21. Log transformed degradation slopes for the *n*-alkane tridecane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

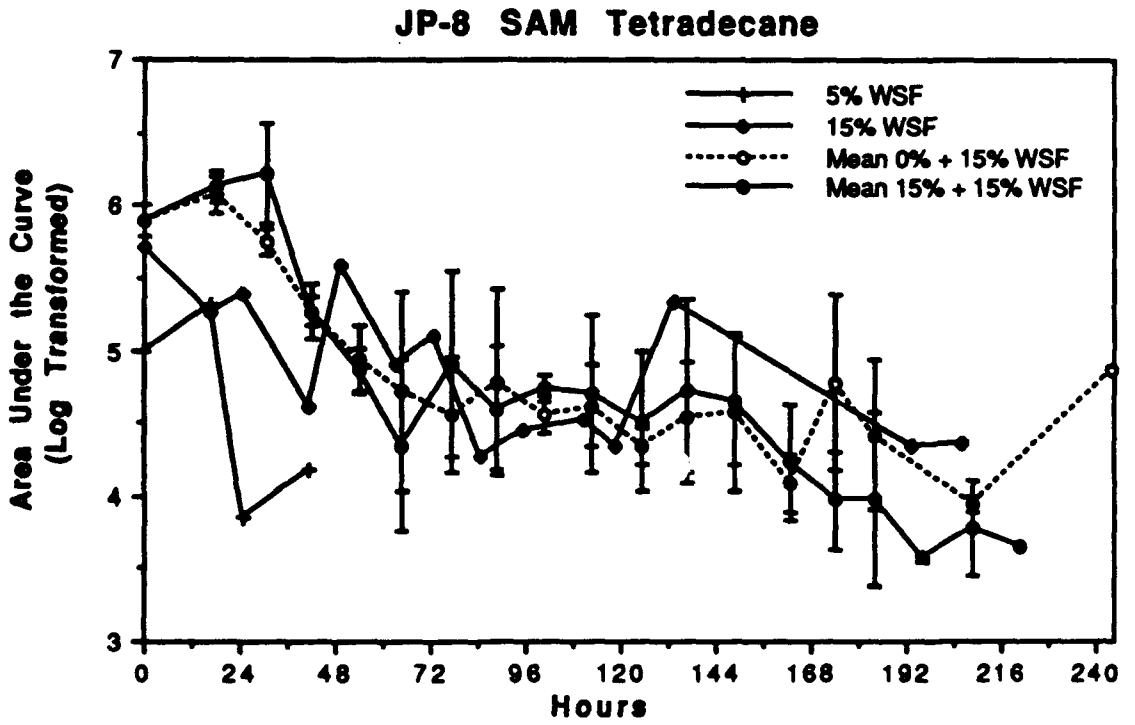


Figure 22. Log transformed degradation slopes for the *n*-alkane tetradecane from the (a) JET-A MFC and (b) JP-8 SAM microcoms.

5% WSF treatment was greater than in any of the other treatment groups. The 0%+15% WSF was second, followed by the 15% WSF, the 1% WSF, and finally the 15%+15% WSF treatment group (Appendix D) (Figure 23a). The 15% WSF degradation rate was not significantly different from the 0%+15% or the 15%+15% WSF treatments; however, the re-treated groups were degraded at significantly different rates compared to each other (Table 11).

The degradation rates of benzene in the 1%, 5%, 15%, and 0%+15% WSF treatments were very similar. The 15% WSF was degraded at the fastest rate, while the 15%+15% WSF treatment was degraded at the slowest rate (Appendix D) (Figure 24a). The 15%, 0%+15%, and 15%+15% WSF treatment groups were degraded at significantly different rates from each other (Table 11).

The mixture of *meta*- and *para*-xylenes were degraded at the fastest rate in the 5% WSF treatment concentration followed by the 15%, 0%+15%, 15%+15%, and the 1% WSF treatments (Appendix D) (Figure 25a). Though there were definite patterns in the rates of degradation for the 15%, 0%+15%, and the 15%+15% WSF treatments, no significant differences were determined between any of these three treatment groups (Table 11).

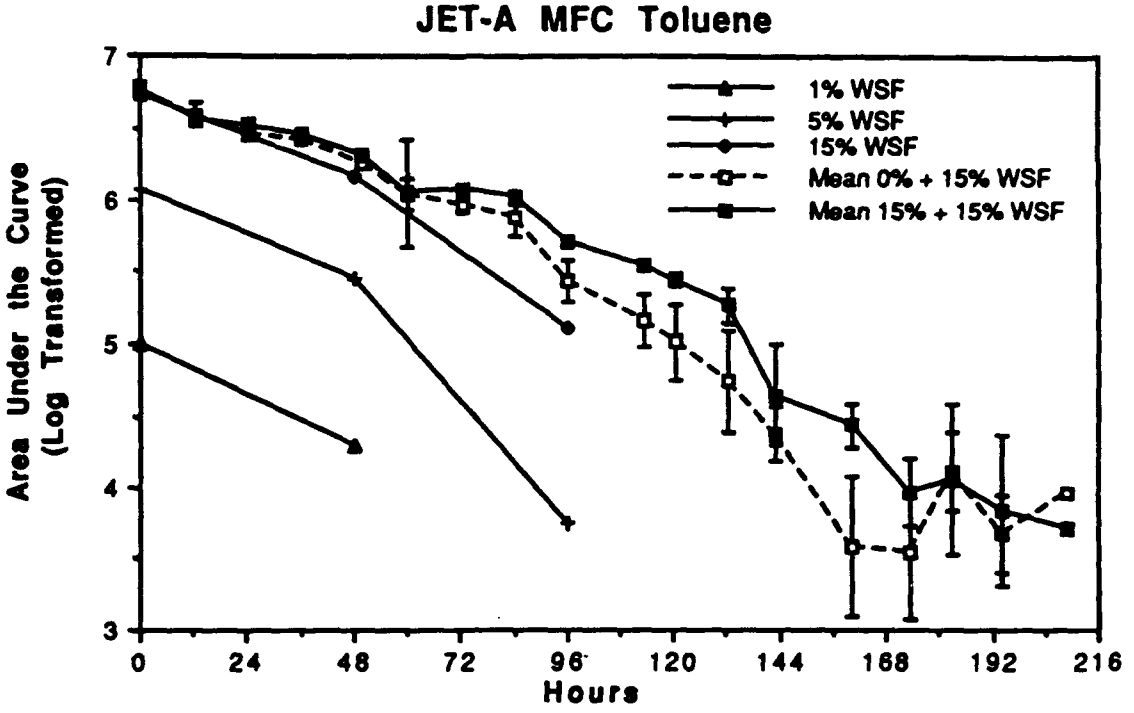
For *ortho*-xylene the patterns in treatment degradation rates were identical to those for the *meta*- and *para*-xylenes. The rank order in degradation rates for the treatment groups were 5%, 15%, 0%+15%, 15%+15%, and the 1% WSF's (Appendix D) (Figure 26a). There were significant differences in the degradation rates between the 15% and the 15%+15% WSF treatments and between the 0%+15% and the 15%+15% WSF treatments (Table 11).

In the alkyl-substituted aromatic group ethylbenzene was consistently degraded at the fastest rate in all of the treatment groups. The highest rate of degradation occurred in the 5% WSF treatment, similar to toluene, and the xylenes. The 15% WSF was second, followed by the 0%+15%, 15%+15%, and finally by the 1% WSF treatment (Appendix D) (Figure 27a). There were no significant differences between any of the three 15% WSF treatment groups (Table 11).

The degradation rate patterns of cyclooctane were similar to benzene, butylbenzene and propylbenzene. The order of degradation was 15%, 0%+15%, 15%+15%, and the 5% WSF being degraded at the slowest rate. There were no significant differences in the rates of degradation between the 15%, 0%+15%, and 15%+15% WSF treatments (Table 12) (Figure 28a).

Butylbenzene and propylbenzene displayed similar degradation rate patterns, that were similar to those for benzene. The fastest rate of degradation occurred in the

a.



b.

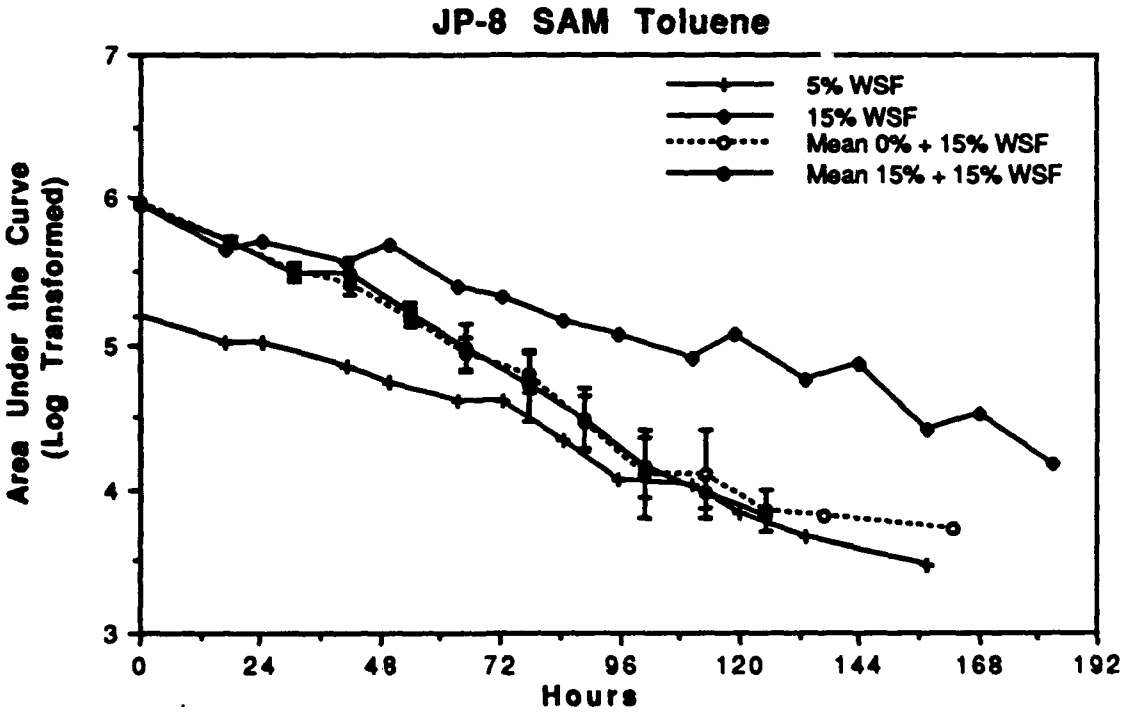
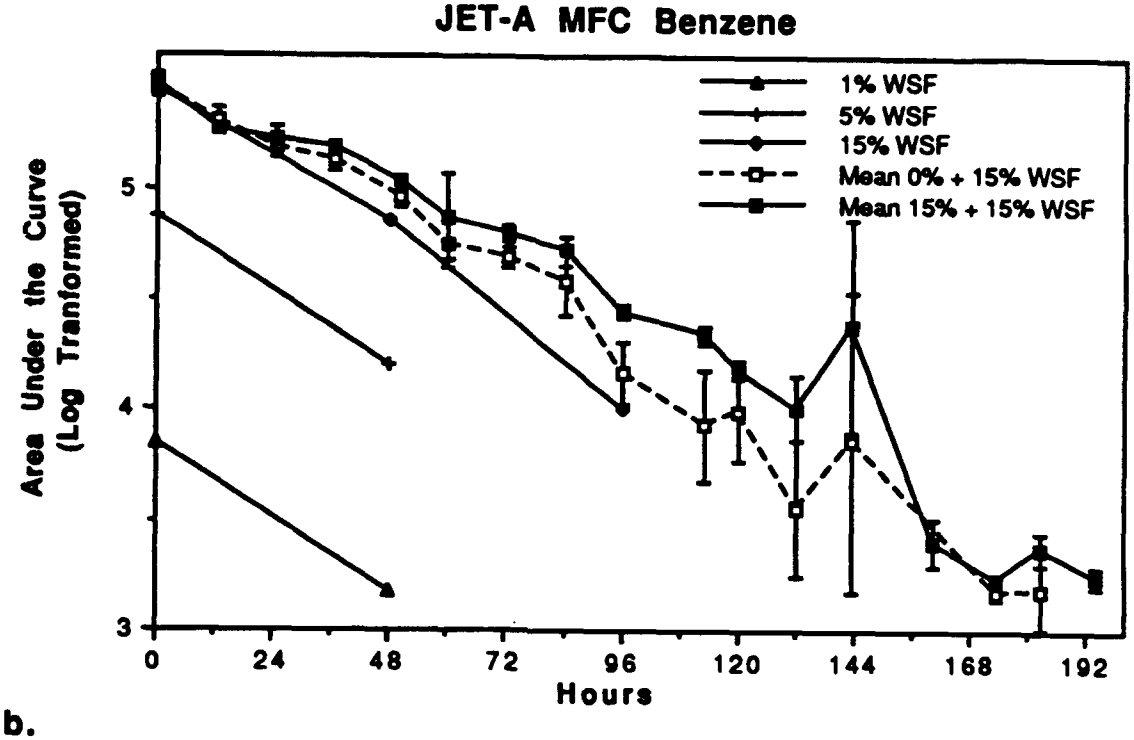


Figure 23. Log transformed degradation slopes for the alkyl-substituted aromatic toluene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

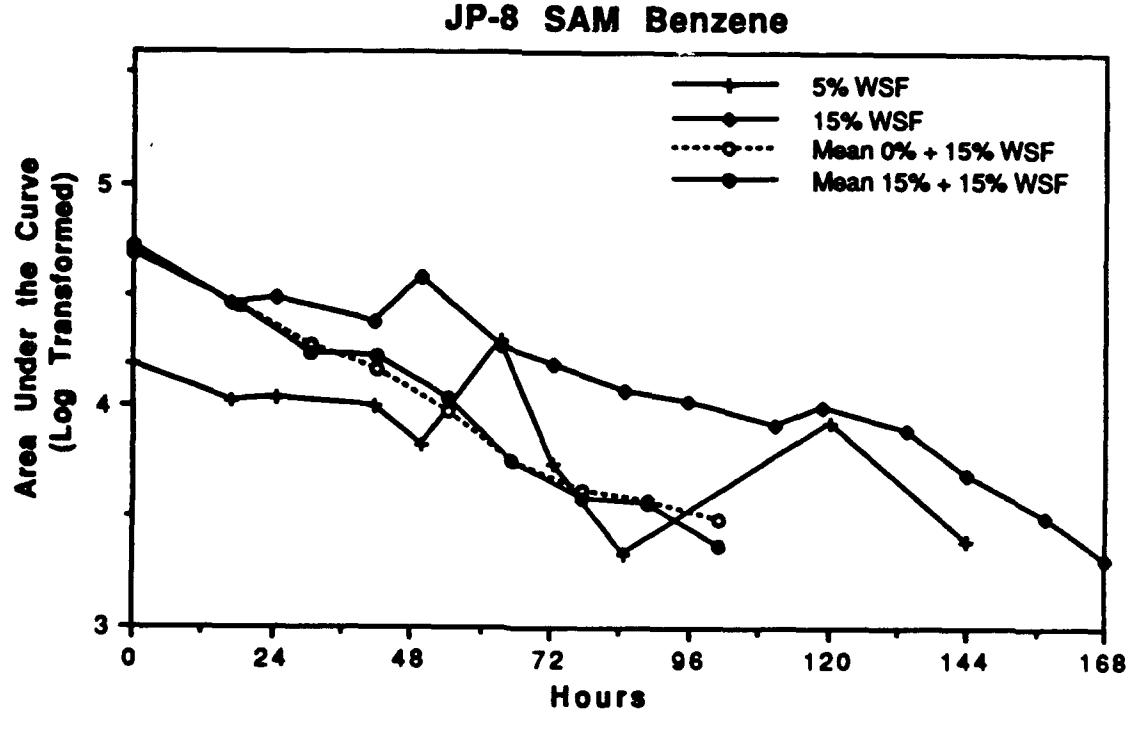
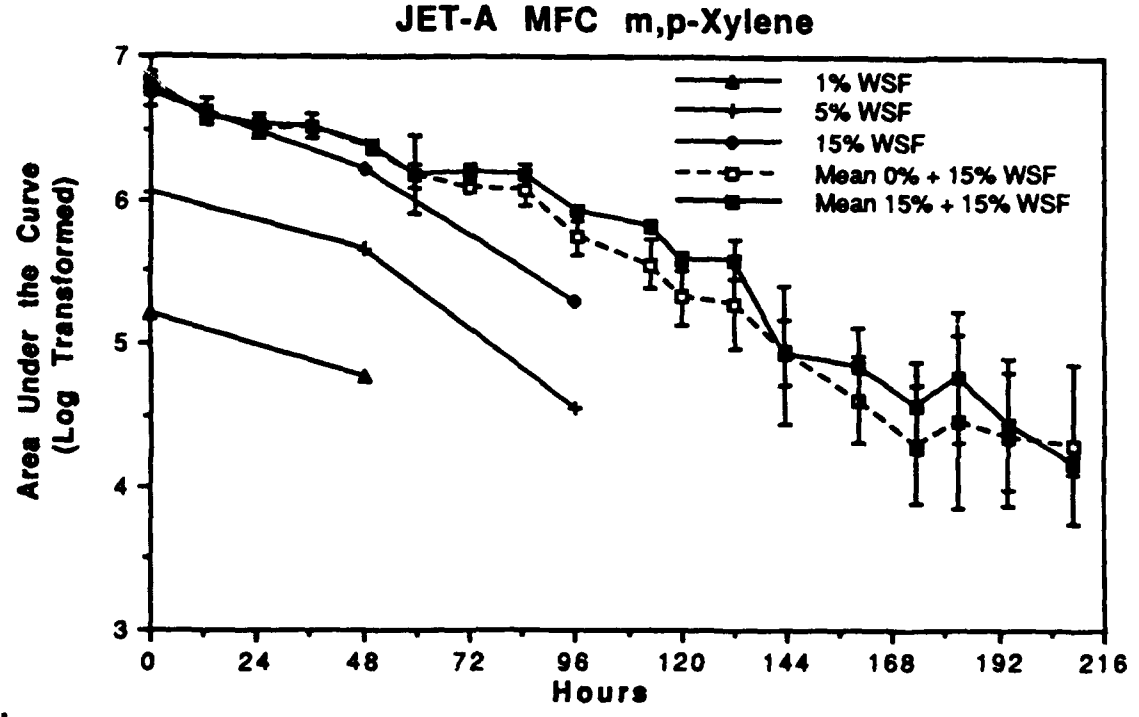


Figure 24. Log transformed degradation slopes for the aromatic hydrocarbon benzene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

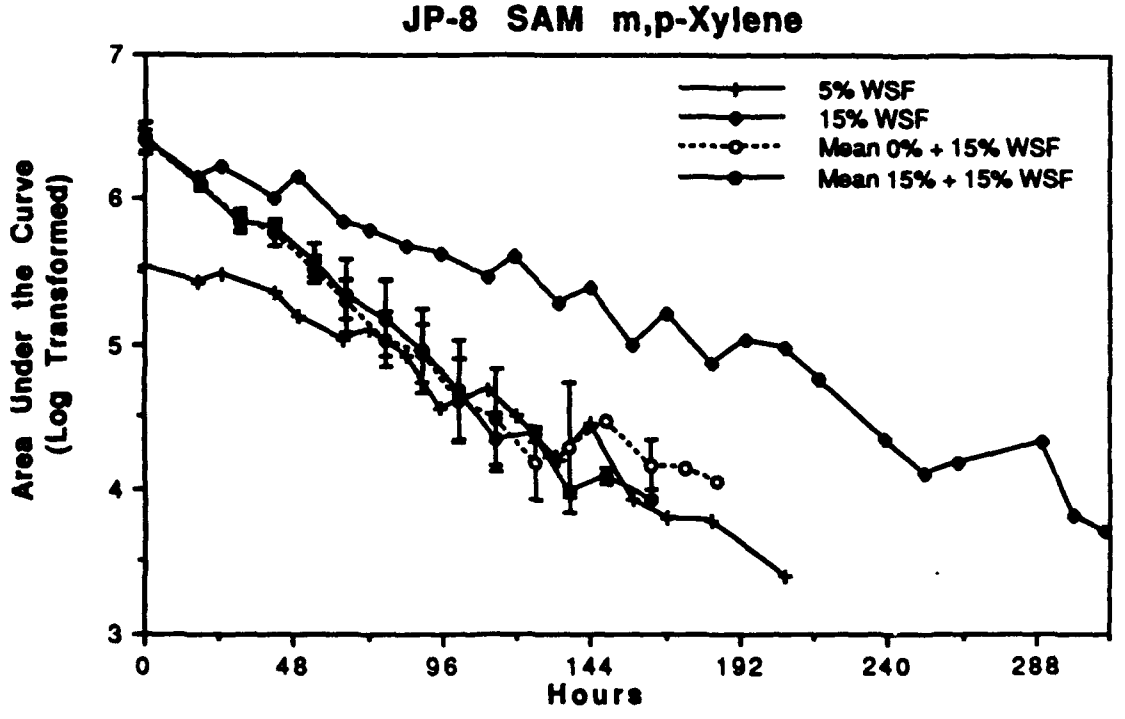
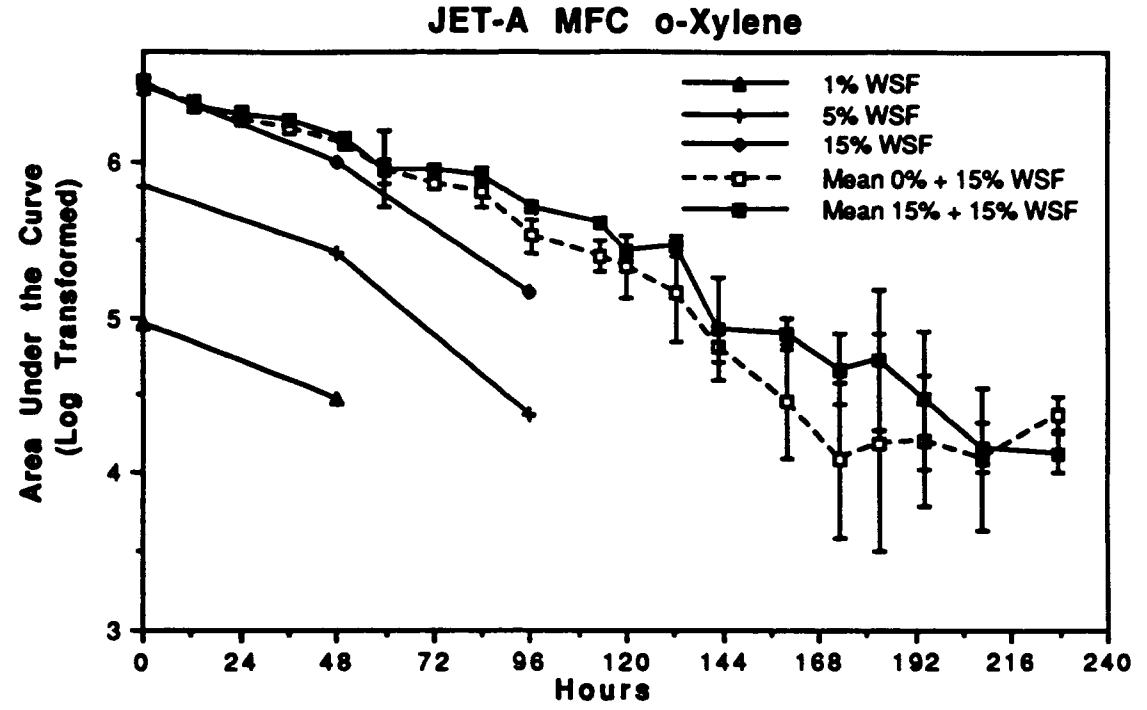


Figure 25. Log transformed degradation slopes for the alkyl-substituted aromatics *meta*- and *para*-xylene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

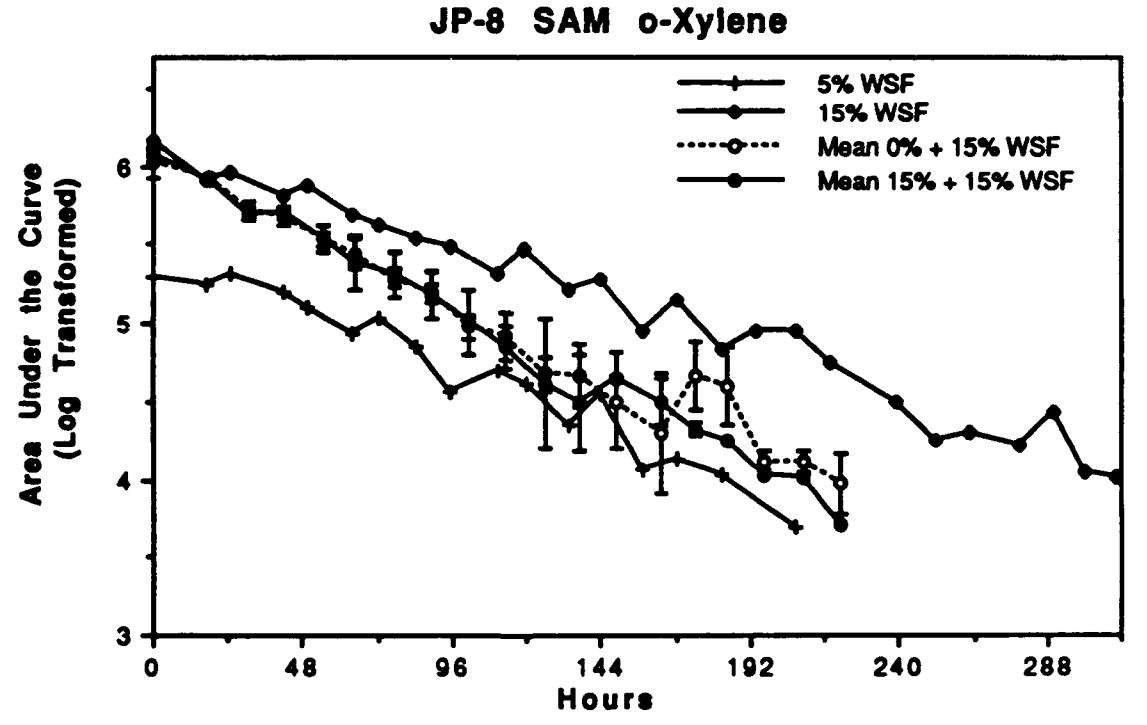


Figure 26. Log transformed degradation slopes for the alkyl-substituted aromatic *ortho*-xylene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

15% WSF treatment, followed by the 5%, the 15%+15%, and the 0%+15% WSF treatments. For butylbenzene there was no significant differences between the two re-treated groups, but they were both significantly different from the 15% WSF treatment group (Table 12) (Figure 29a). For propylbenzene there was no significant difference between the 15% WSF and the 15%+15% WSF treatments, or between the 0%+15% and the 15%+15% WSF groups. However, the 15% WSF was significantly different from the 0%+15% WSF treatment (Table 12) (Figure 30a).

The degradation rates for many of the hydrocarbon components in the 0%+15% and 15%+15% WSF treatments were significantly different from the 15% WSF treatments. The degradation rates in the 15% WSF treatments were consistently faster than the rates of degradation in the 0%+15% and 15%+15% WSF treatments (Figures 19-30) (Tables 10-12). A summary of the combined hydrocarbon components ranked by their rates of degradation is listed (Table 13).

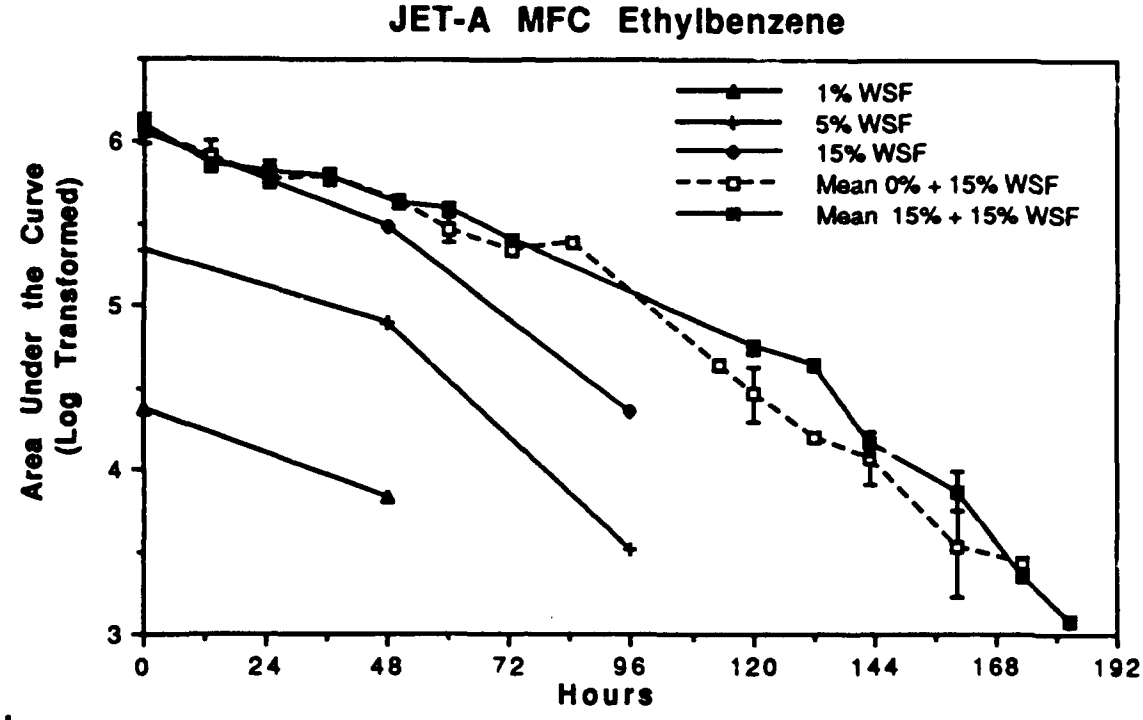
The hydrocarbon metabolites produced during the course of the experiment in the treatment groups, fluctuated in concentration that was dependent on their rates of utilization by the microorganisms and on the rates of the degradation of the parent compounds. Propane, 2-methylpropane, 2,4-dimethylpentane, *trans*-2-pentene, and hexane concentrations were substantial increased at the forty-eight hour, one hundred forty-four hour and at approximately the two hundred and sixteen hour time periods during the course of the experiment (Figures 31a, 34a, 35a, 37a, and 38a). *cis*-2-Pentene was also highly variable in concentration (Figure 32a). Only 3-methylpentane remained relatively constant throughout the experiment (Figure 36a).

JP-8 SAM Results

The alkanes were the major components in the water soluble fraction of JP-8 (Figures 17b - 18b) (Table 3). Dodecane was the highest in concentration in all of the WSF treatment groups, as in the JET-A MFC. Tridecane was the second highest in concentration level with decane and tetradecane being third and fourth, respectively (Tables 3 and 4b). The alkyl-substituted aromatics propylbenzene and butylbenzene were next in concentration with ethylbenzene and cyclooctane at much lower levels. The aromatics *m*,*p*-xylenes, toluene, *o*-xylene, and benzene, were present in the lowest concentrations and were much lower in comparison to the JET-A water soluble fraction concentrations (Tables 2-3).

These hydrocarbon components were also ranked by concentration levels in their respective hydrocarbon classes (Tables 4b, 6b, and 8b). Only the aromatics maintained the same concentration rankings in each of the treatment groups (Table 6b). In the

a.



b.

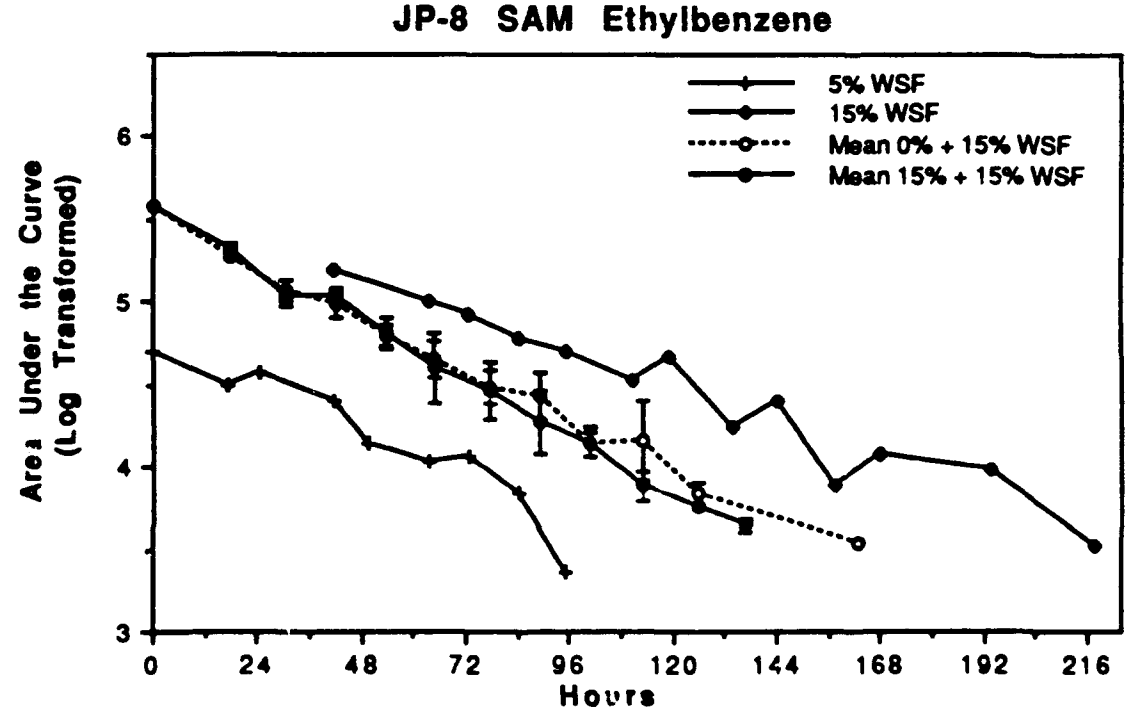
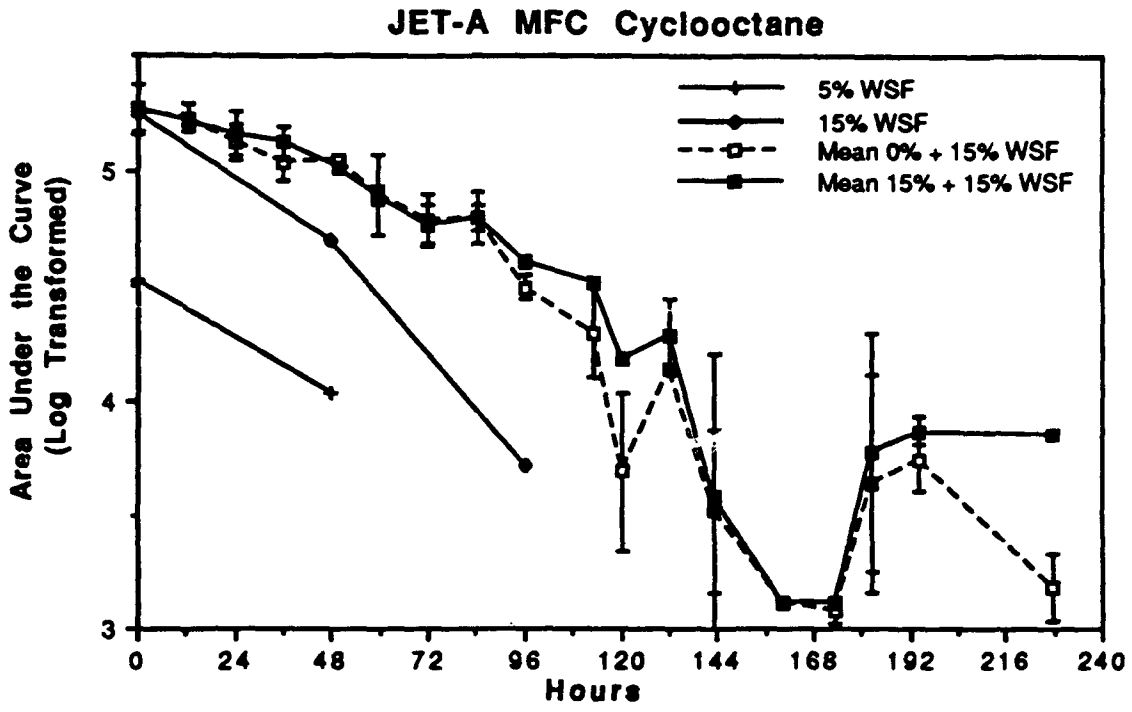


Figure 27. Log transformed degradation slopes for the alkyl-substituted aromatic ethylbenzene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

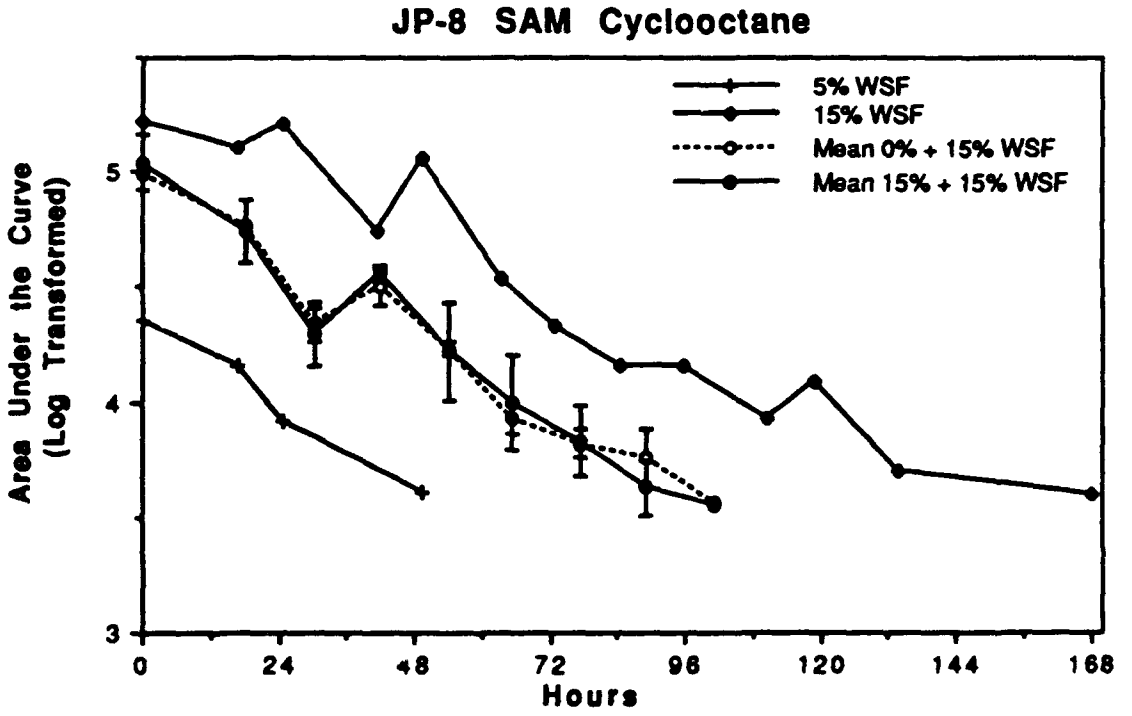
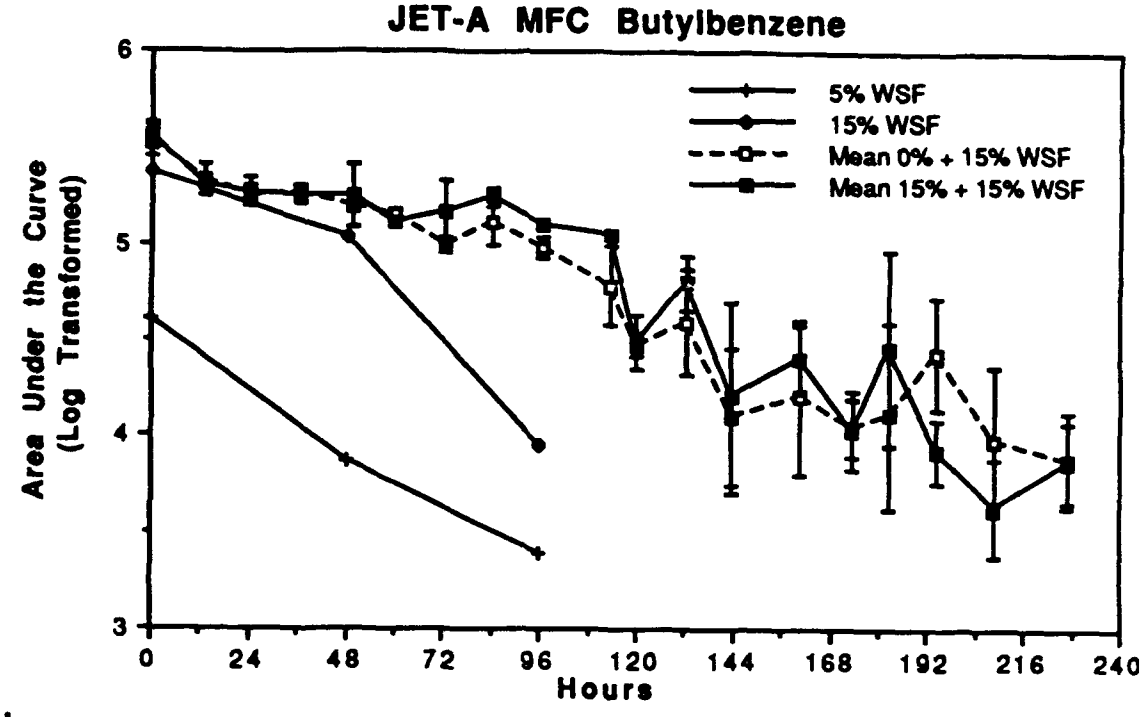


Figure 28. Log transformed degradation slopes for the cyclo-alkane cyclooctane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

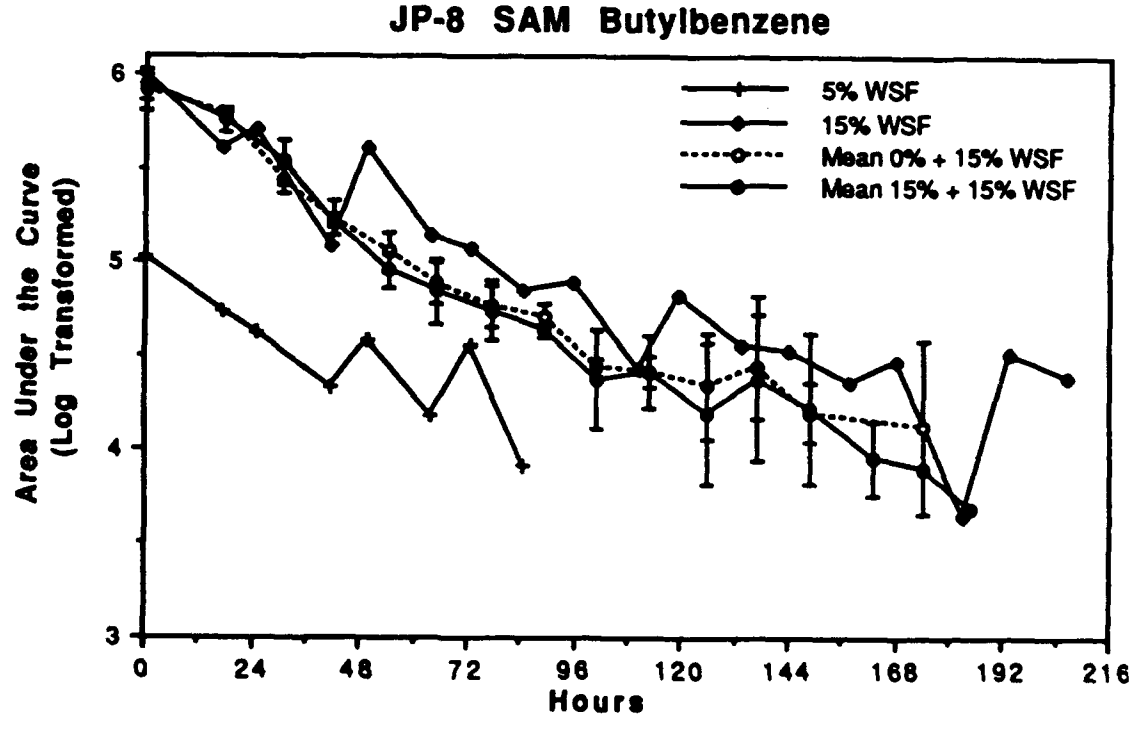
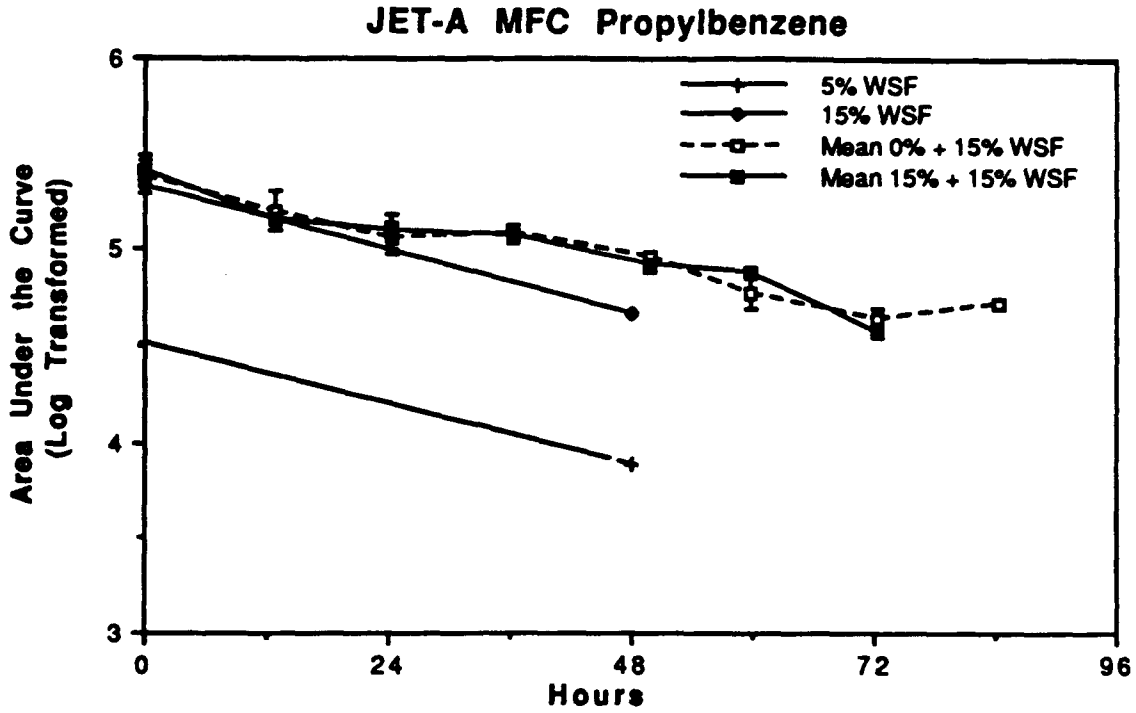


Figure 29. Log transformed degradation slopes for the alkyl-substituted aromatic butylbenzene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

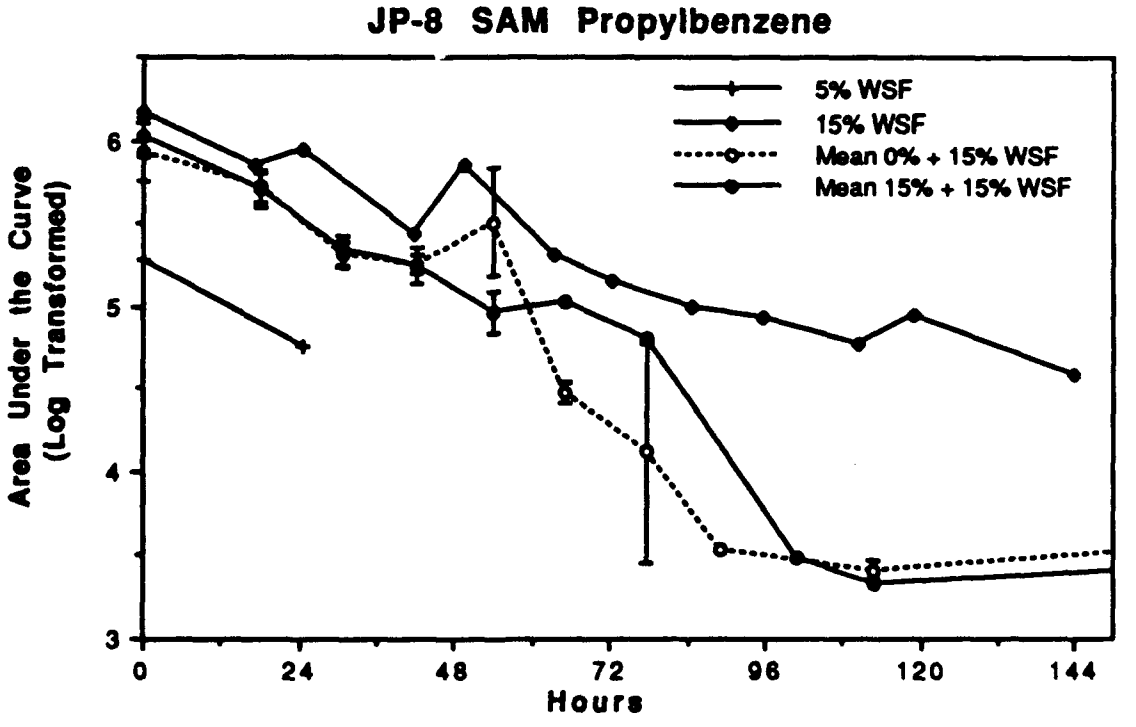
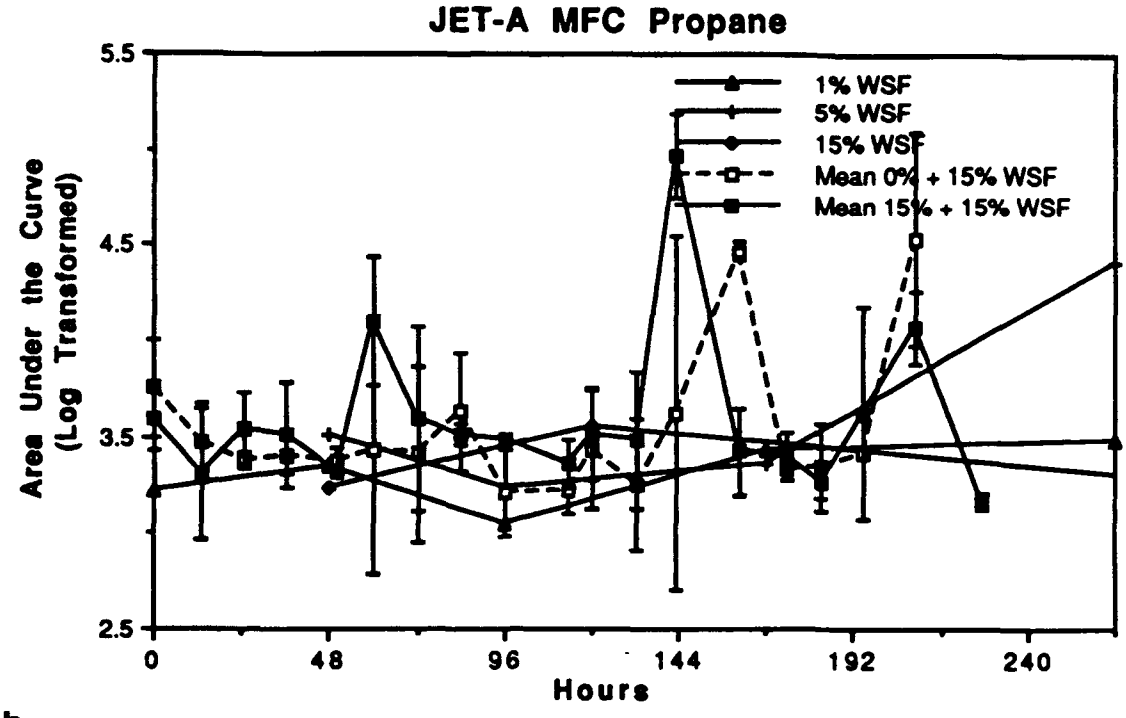


Figure 30. Log transformed degradation slopes for the alkyl-subtituted aromatic propylbenzene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

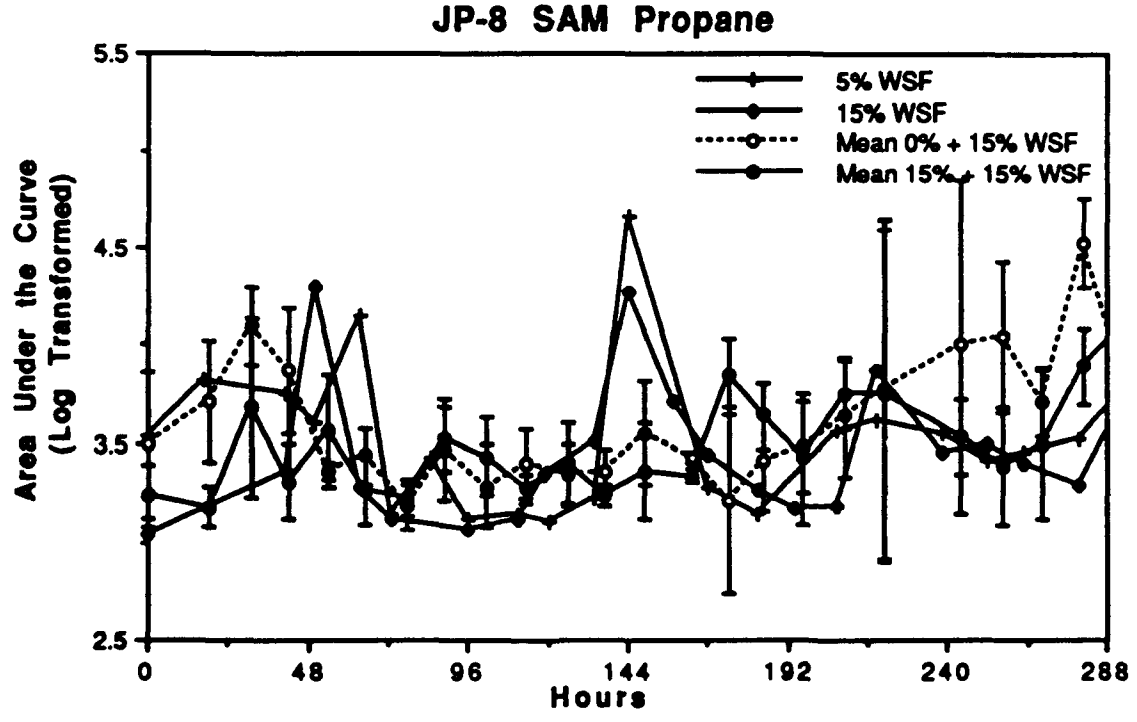
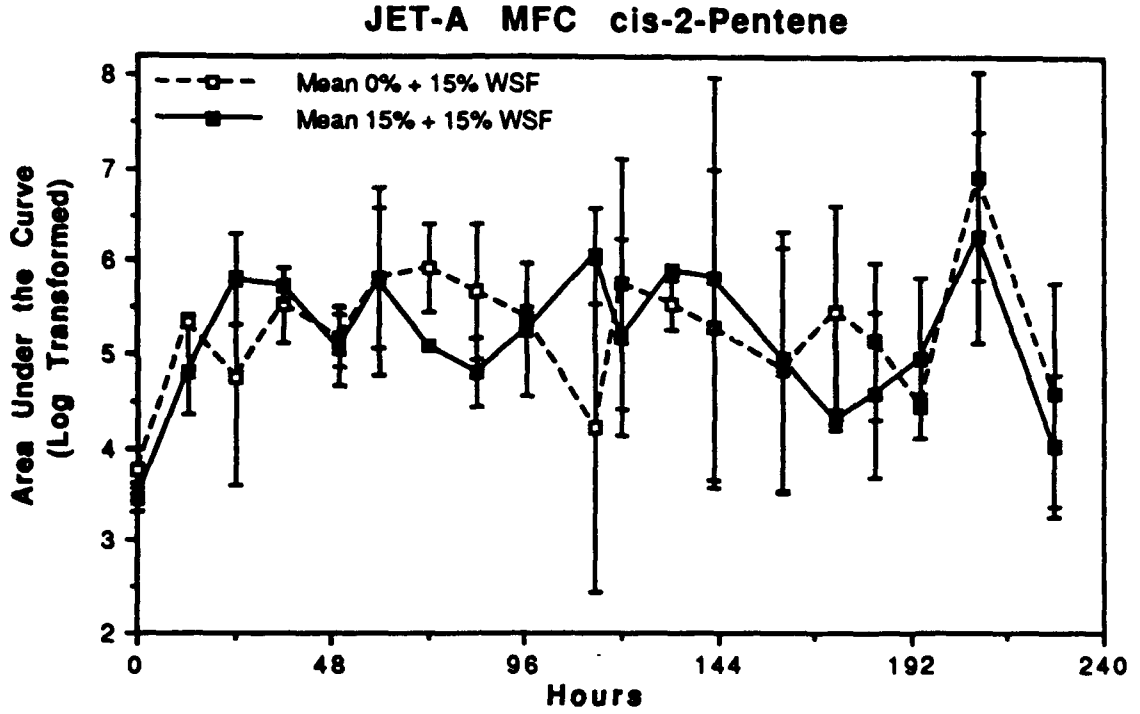


Figure 31. Log transformed data for the *n*-alkane metabolic by-product propane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

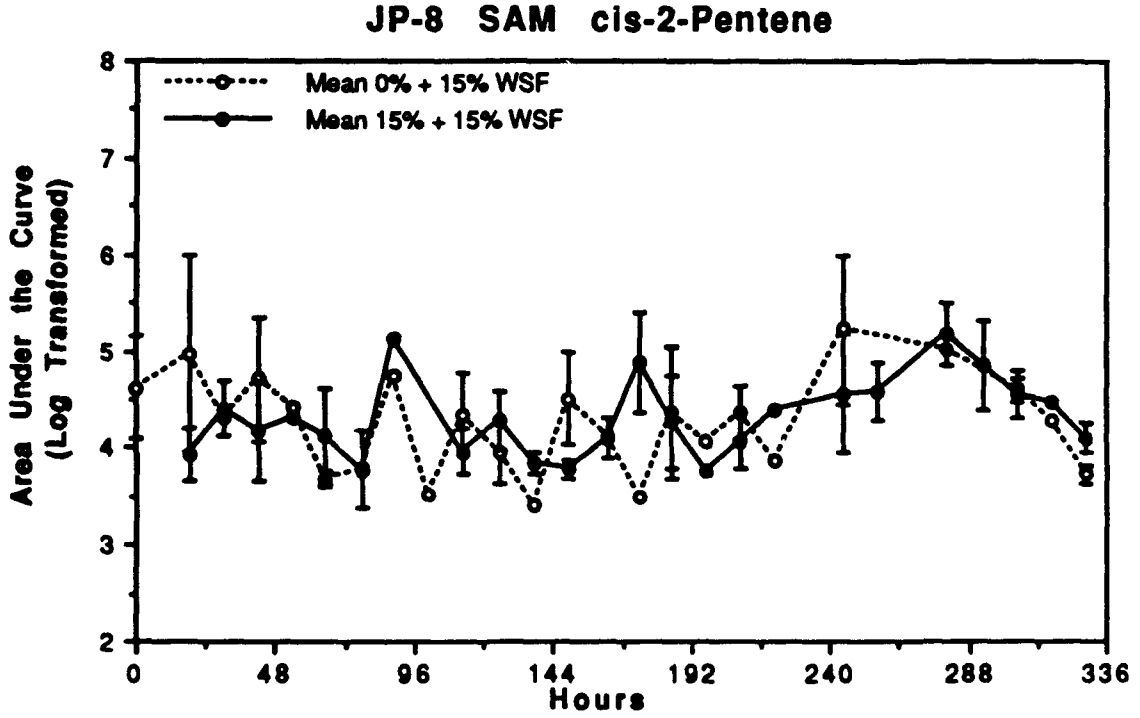
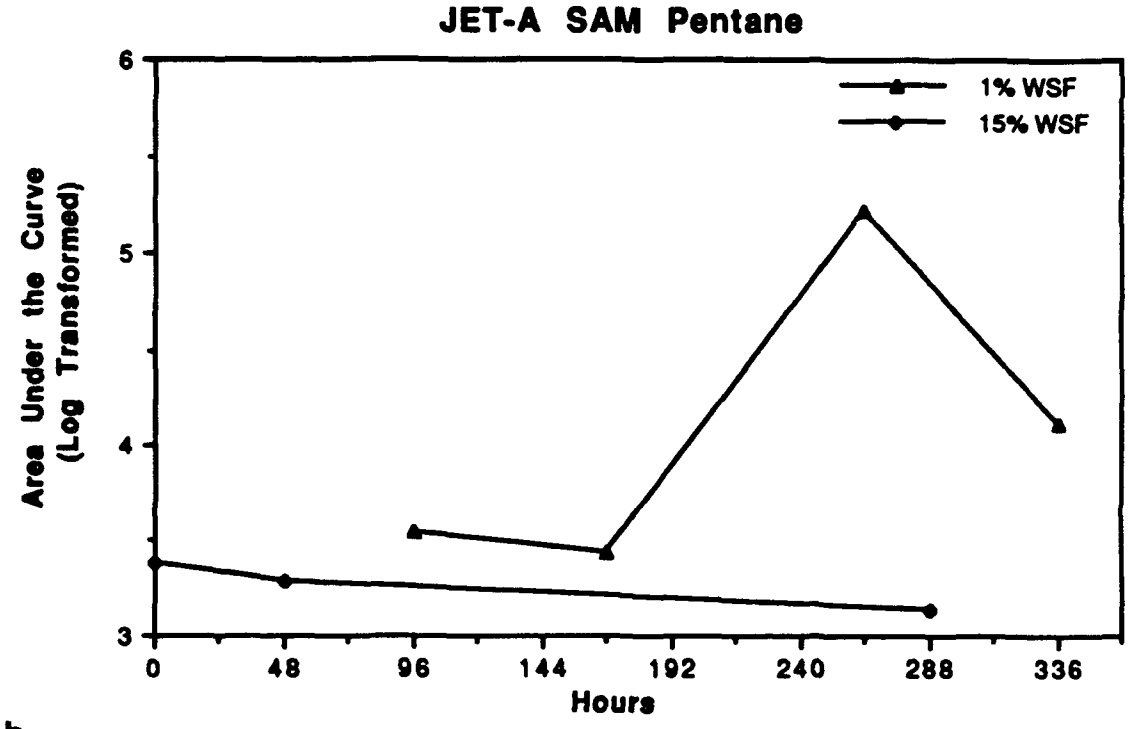


Figure 32. Log transformed data for the alkene metabolic by-product *cis*-2-pentene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

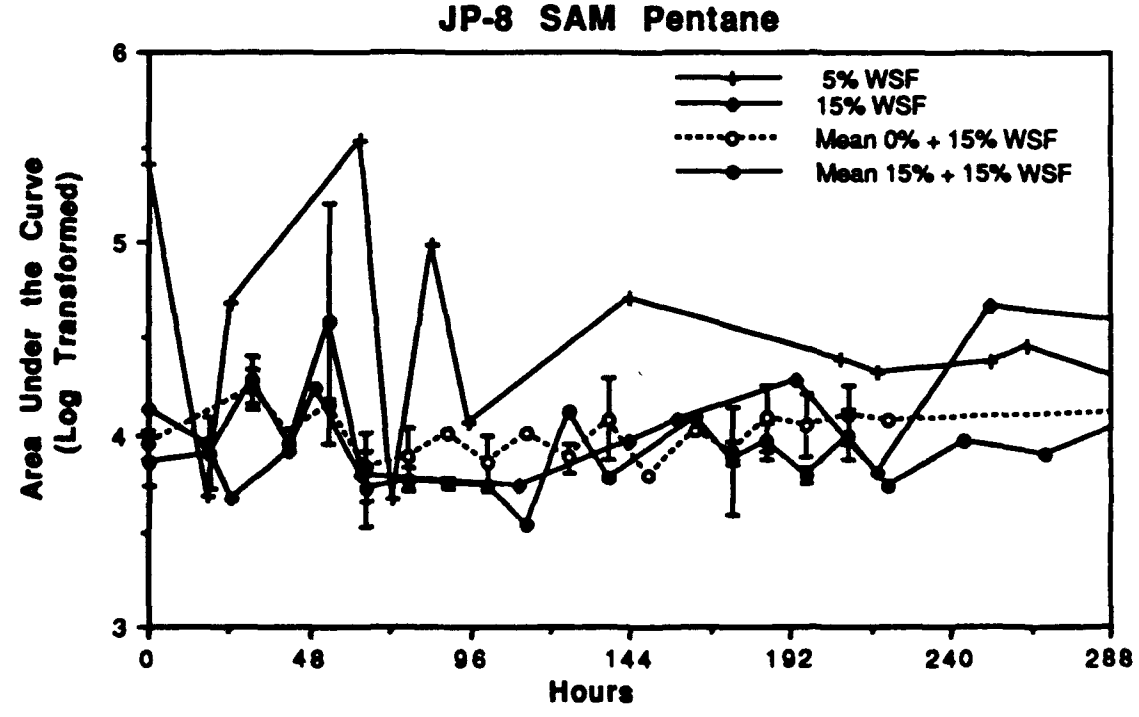
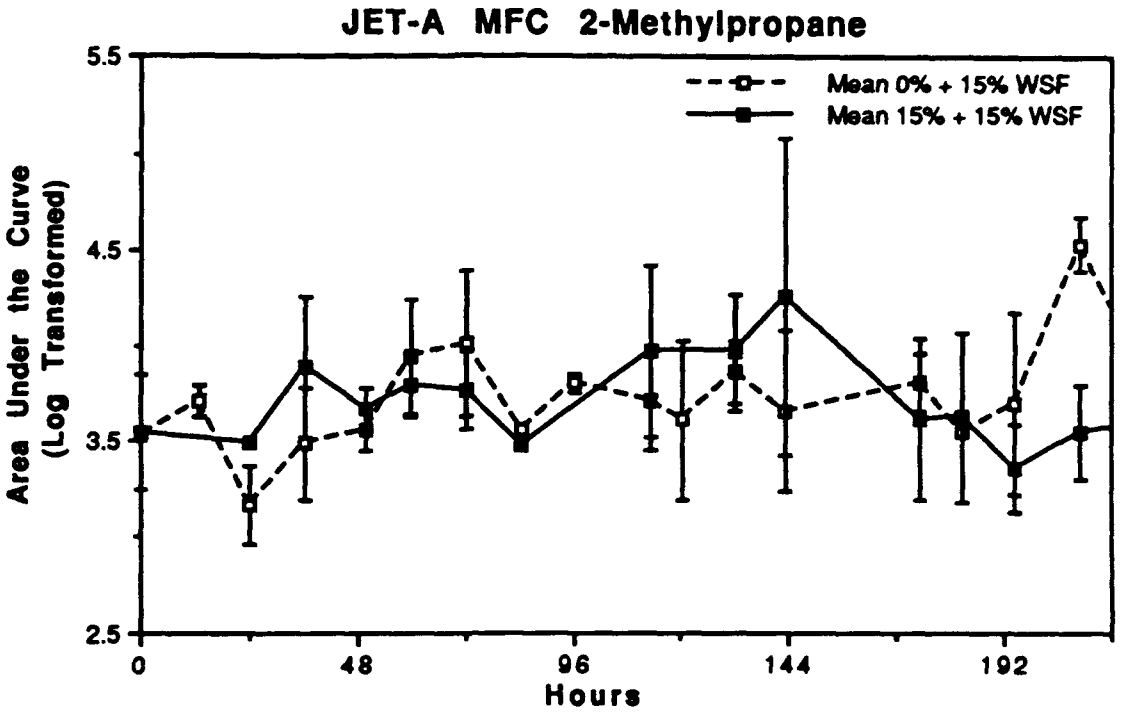


Figure 33. Log transformed data for the *n*-alkane metabolic by-product pentane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

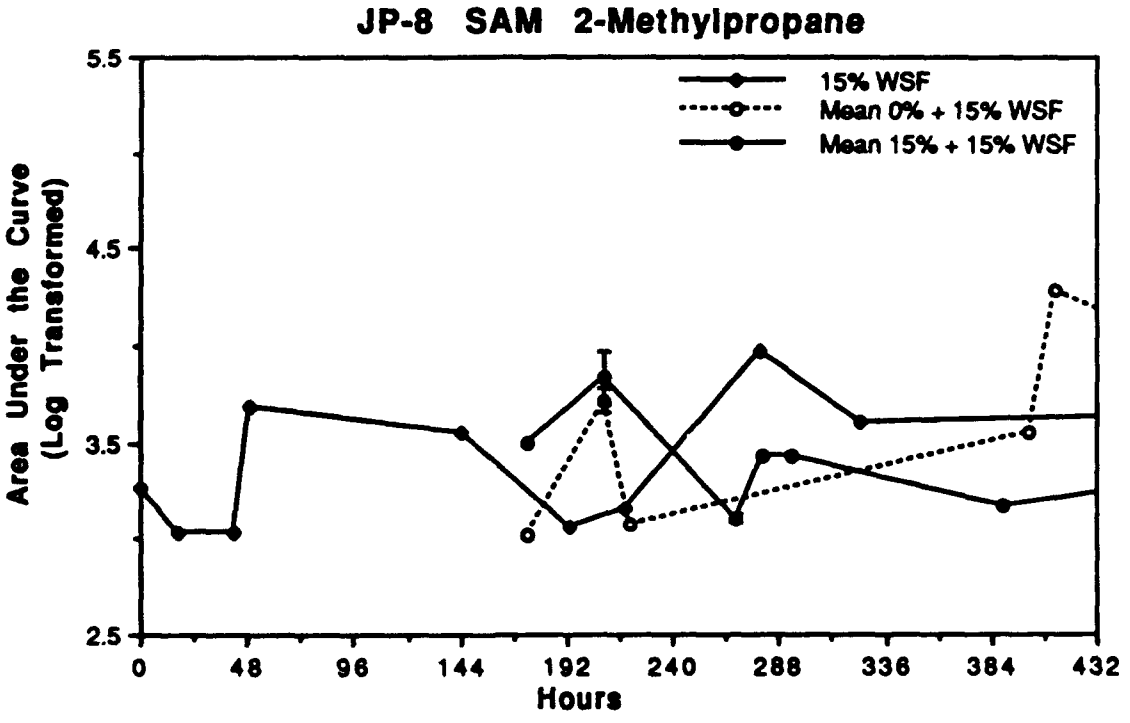
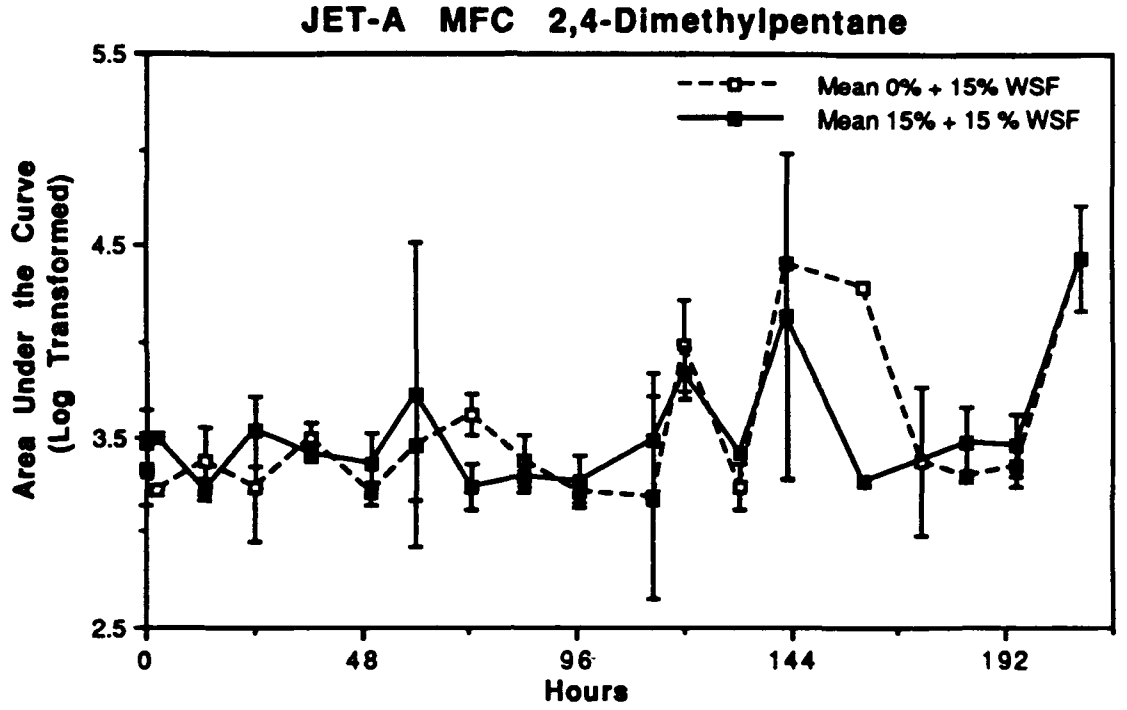


Figure 34. Log transformed data for the branched alkane metabolic by-product 2-methylpropane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

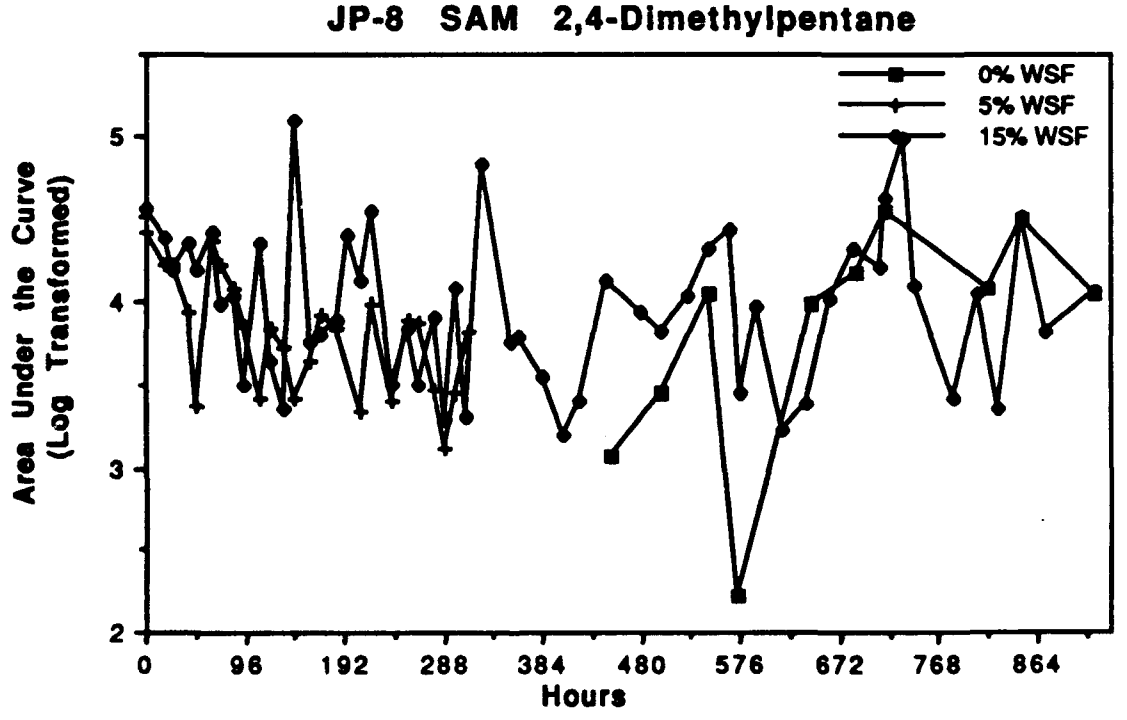
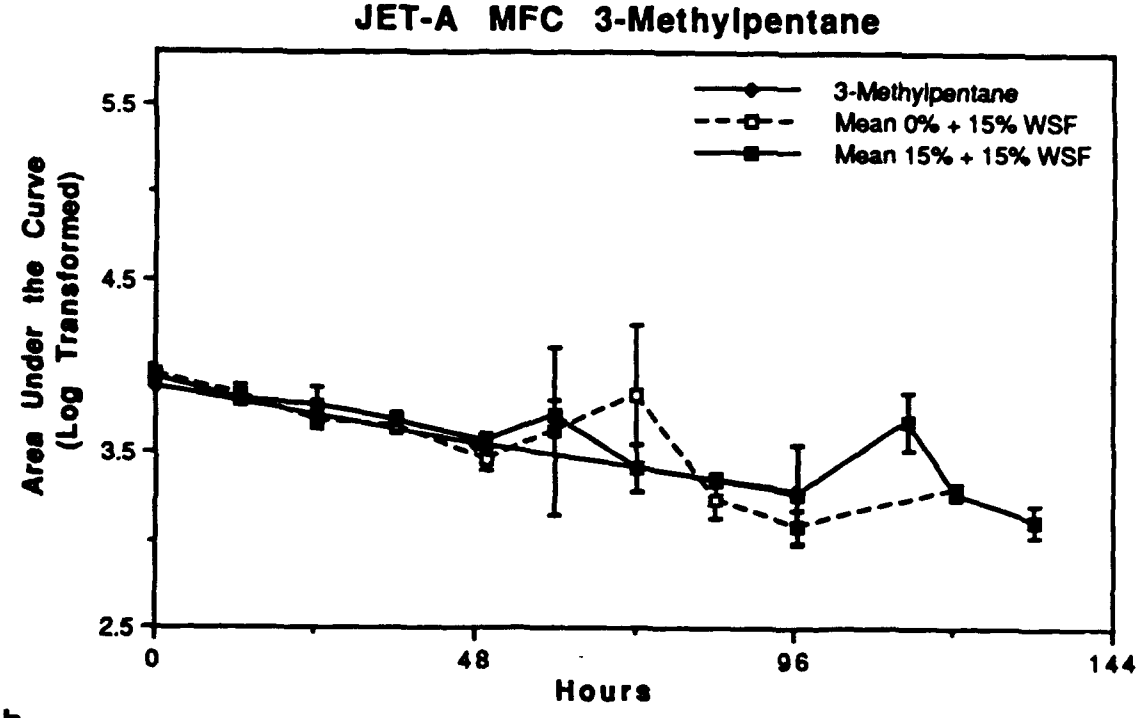


Figure 35. Log transformed data for the branched alkane metabolic by-product 2,4-dimethylpentane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

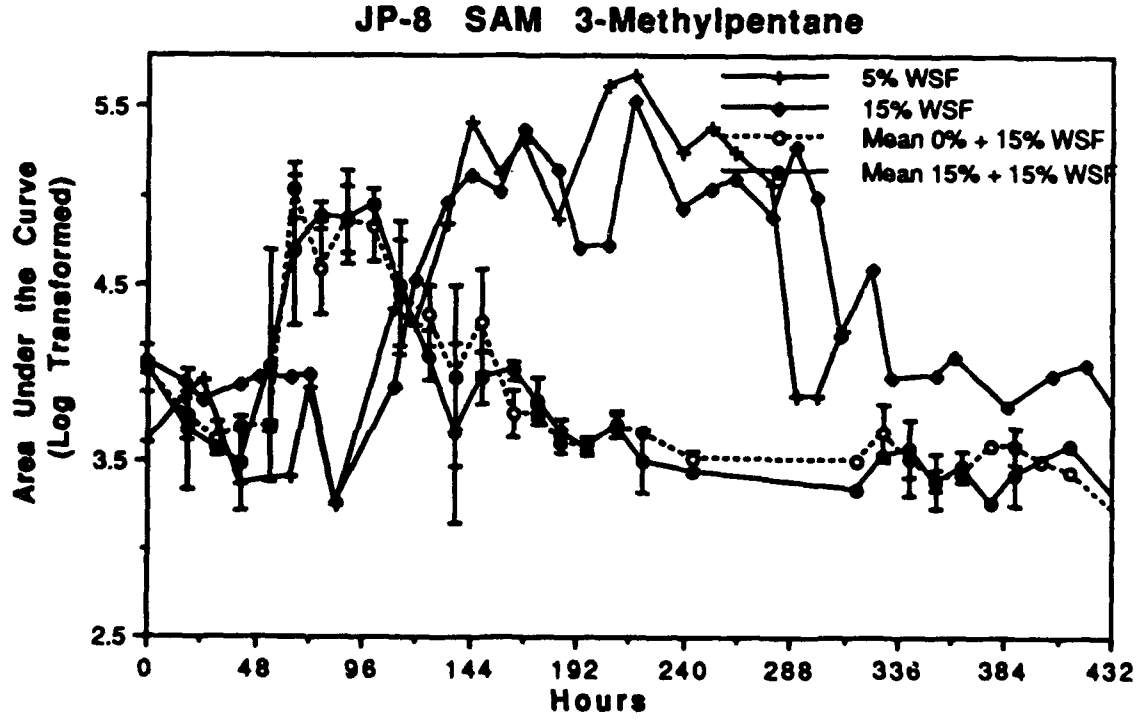
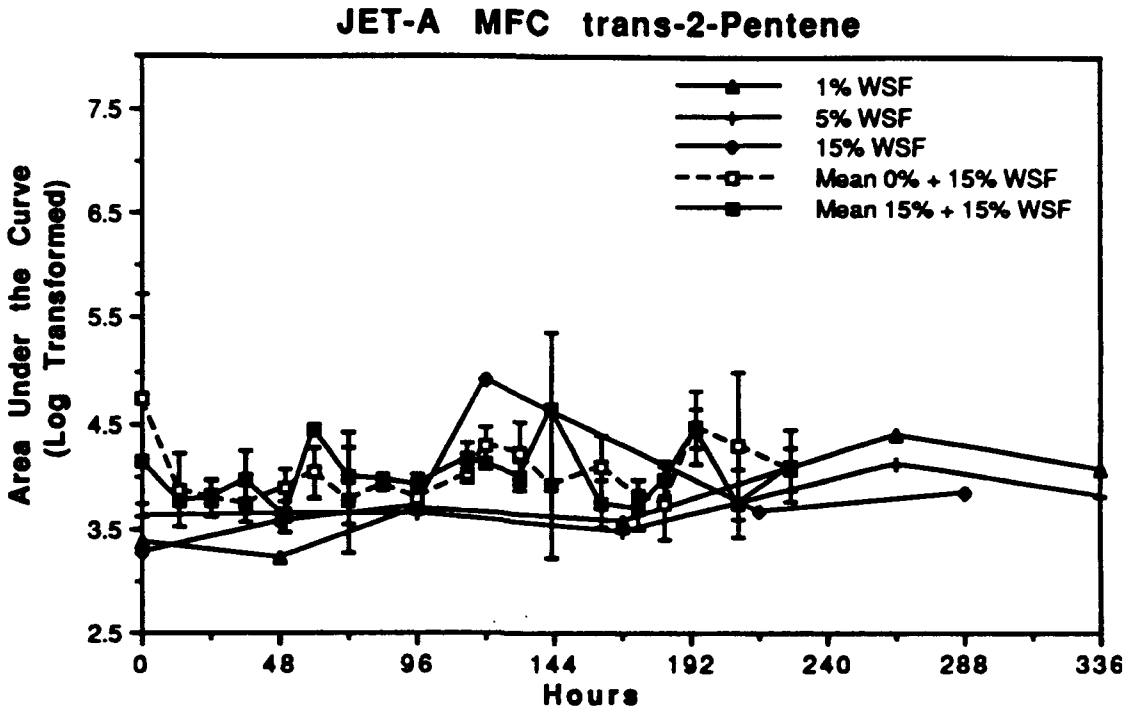


Figure 36. Log transformed data for the branched alkane metabolic by-product 3-methylpentane, from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

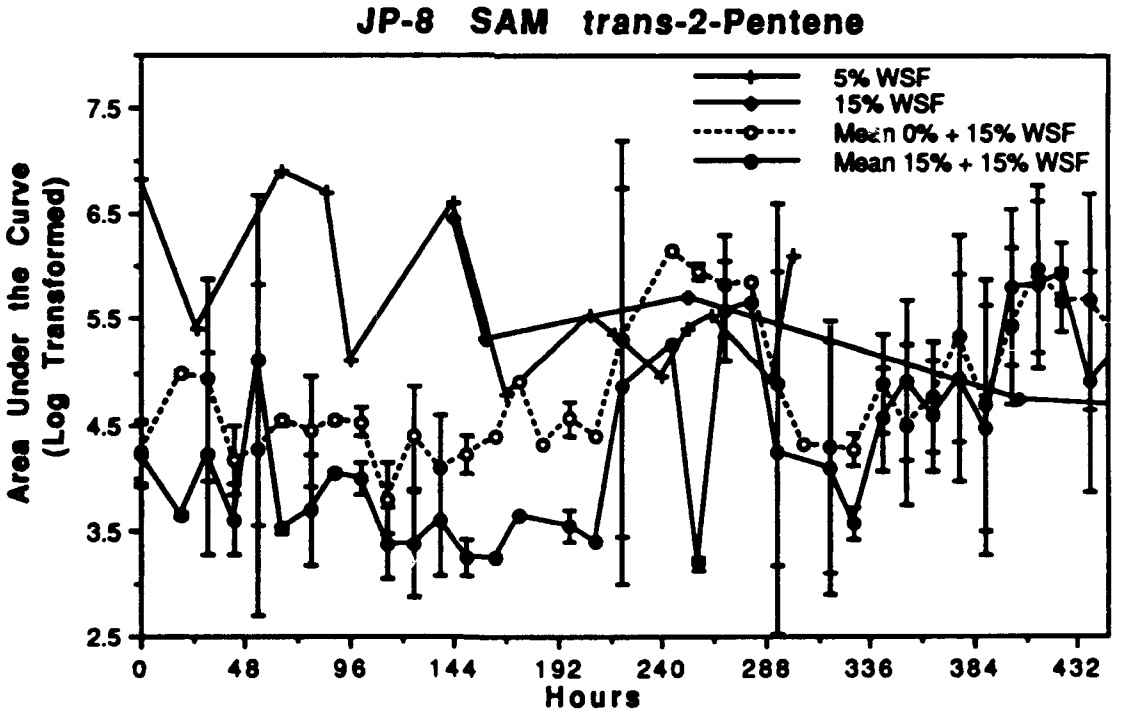
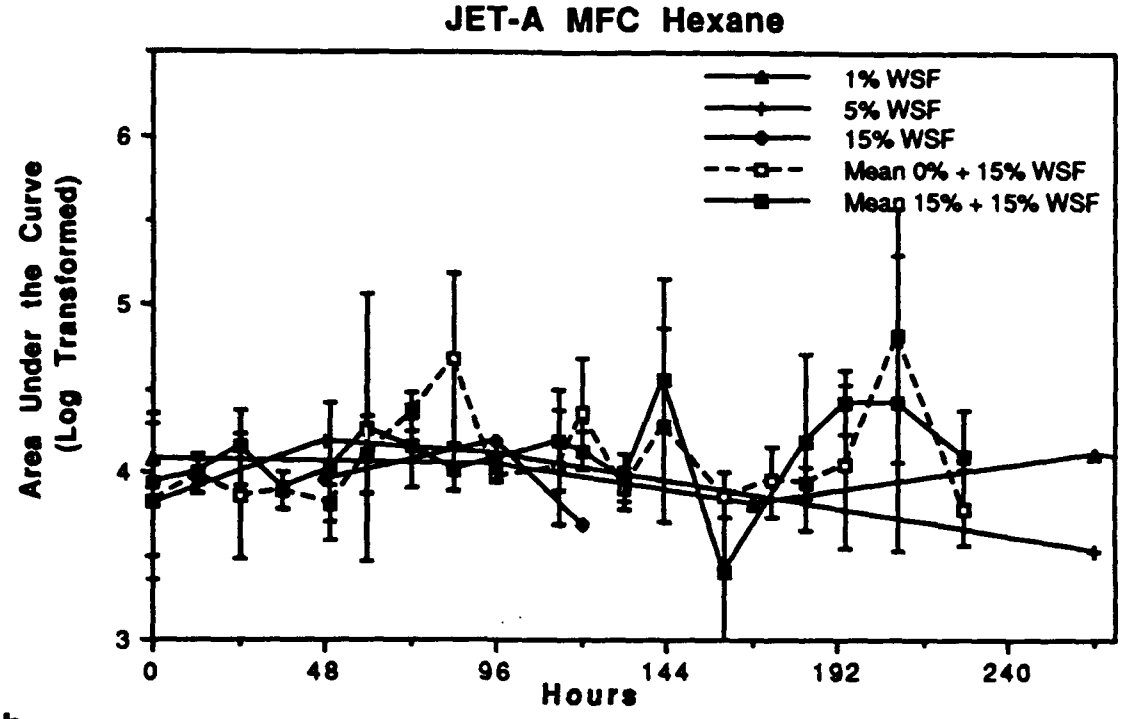


Figure 37. Log transformed data for the alkene metabolic by-product *trans*-2-pentene from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

a.



b.

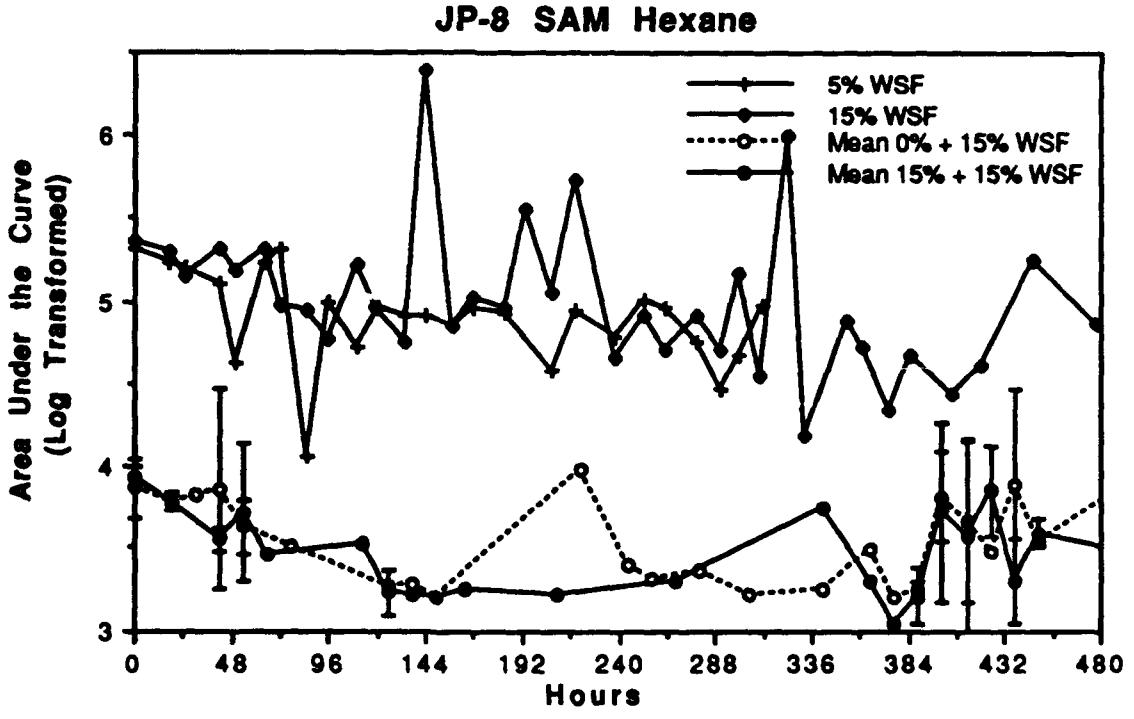


Figure 38. Log transformed data for the *n*-alkane metabolic by-product hexane from the (a) JET-A MFC and (b) JP-8 SAM microcosms.

Table 13. Summary of the ranked order of degradation rates for all JET-A MFC hydrocarbon components in all treatment groups.

1% WSF	5% WSF	15% WSF	0%+15% WSF	15%+15% WSF
Decane	Toluene	Decane	Tetradecane	Toluene
Toluene	Ethylbenzene	Ethylbenzene	Toluene	Ethylbenzene
Benzene	Dodecane	Toluene	Tridecane	Dodecane
Ethylbenzene	*m,p*-Xylene	Cyclooctane	Ethylbenzene	Cyclooctane
o-Xylene	*o*-Xylene	Tridecane	Dodecane	*m,p*-Xylene
m,p-Xylene	Decane	Benzene	Benzene	Tetradecane
	Benzene	*m,p*-Xylene	*m,p*-Xylene	Tridecane
	Propylbenzene	Butylbenzene	Cyclooctane	Benzene
	Butylbenzene	Propylbenzene	*o*-Xylene	Decane
	Cyclooctane	*o*-Xylene	Decane	*o*-Xylene
		Dodecane	Propylbenzene	Propylbenzene
			Butylbenzene	Butylbenzene

Table 14. Ranked order of degradation for all JP-8 SAM hydrocarbon components in all treatment groups.

1% WSF	5% WSF	15% WSF	0%+15% WSF	15%+15% WSF
Ethylbenzene	Tetradecane	Tetradecane	Propylbenzene	Tetradecane
	Tridecane	Propylbenzene	Tetradecane	Propylbenzene
	Propylbenzene	Butylbenzene	Tridecane	Tridecane
	Cyclooctane	Cyclooctane	Toluene	Toluene
	Toluene	Ethylbenzene	Cyclooctane	*m,p*-Xylene
	m,p-Xylene	Toluene	*m,p*-Xylene	Cyclooctane
	Ethylbenzene	*m,p*-Xylene	Ethylbenzene	Ethylbenzene
	Butylbenzene	Tridecane	Benzene	Benzene
	o-Xylene	Propylbenzene	Decane	Decane
	Benzene	Benzene	Butylbenzene	Butylbenzene
	Dodecane	*o*-Xylene	*o*-Xylene	*o*-Xylene
	Decane	Decane	Dodecane	Dodecane
		Dodecane		

Table 3. Initial concentrations (µg/L) of the individual hydrocarbon components in the JP-8 SAM water soluble fraction treatment groups.

	JP-8 SAM Treatment Group Concentrations (µg/L)					
Hydrocarbon	0% WSF	1% WSF	5% WSF	15% WSF	0%+15% WSF	15%+15% WSF
Dodecane			534.5	3988.7	3474.8	4025.3
Tridecane			377.7	2685.1	2569.3	3124.7
Decane		35.5	393.4	2317.0	1984.0	1919.0
Tetradecane			299.0	1581.0	2493.4	2455.8
m,p-Xylene		6.2	159.6	1211.3	1281.9	1278.9
Propylbenzene			122.8	966.3	585.3	699.2
Butylbenzene			95.9	893.3	736.6	776.6
Toluene		8.5	97.1	541.1	579.3	558.1
o-Xylene		4.0	76.0	535.5	395.8	452.2
Butane	0.2	130.8	250.4	371.9		
Hexane	5.5	78.5	163.3	183.9	6.1	7.0
Ethylbenzene			54.6	168.0	411.9	398.6
Cyclooctane			14.5	104.8	63.1	71.4
Benzene		16.1	29.9	101.8	97.2	93.7
trans-2-Pentene			12520.7	5436.3	85.4	38.0
Pentane		57.9	5083.2	270.4	181.7	148.8
Octane				55.2	52.5	52.9
2,4-Dimethylpentane	1.2	10.1	25.7	34.5		
2-Methylpentane		8.3	10.8	38.6		
Propane	10.0	41.0	94.3	29.5	105.9	48.9
3-Methylpentane	3.5	2.3	3.8	11.1	10.7	10.4
2-Methylpropane		1.1		1.4		
cis-2-Pentene					122.8	

alkanes dodecane was the only hydrocarbon component to have the same ranking in all treatment groups. The 0%+15% and the 15%+15% WSF treatments had the same rank order of hydrocarbon component concentrations, with dodecane first followed by tridecane, tetradecane, and decane. The rank order was not consistent when compared to the other treatment groups (Table 4b). For the alkyl-substituted aromatics, propylbenzene was highest in the 5% and the 15% WSF treatments, with butylbenzene, ethylbenzene, and cyclooctane following (Table 8b). In the 0%+15% and the 15%+15% WSF treatments propylbenzene and butylbenzene were reversed with butylbenzene being highest, followed by propylbenzene, ethylbenzene, and cyclooctane. Unlike the JET-A MFC water soluble component concentrations, the concentrations of hydrocarbon components in the JP-8 SAM were not consistent in their rank order of concentrations in each of the treatment groups, except for the aromatics.

The concentrations of the individual hydrocarbon components increased as the percentage of the WSF treatment increased. The exception was decane, which decreased in concentration as the jet fuel water soluble fraction was increased (Table 4b). In the 0%+15% and the 15%+15% WSF treatments some components were also less in concentration than in the original 15% WSF treatment. Some of the compounds that were less concentrated were benzene, butylbenzene, cyclooctane, decane, propylbenzene, and *o*-xylene. Some compounds that were slightly elevated in the 0%+15% WSF treatments, but lower in the 15%+15% WSF treatments were decane, ethylbenzene, tetradecane, toluene, and *m,p*-xylene. The only substantial increases in hydrocarbon component concentrations above the 15% WSF treatment were for ethylbenzene that was three times greater and for tetradecane that was one and one half times higher (Table 3).

As in the JET-A MFC experiment, the increase in the component concentration levels were not consistent with the concentrations of the WSF amendments (Table 3). A comparison of the individual hydrocarbon concentrations in the 15% WSF to the 5% WSF concentrations showed that most components increased seven fold. The 0%+15% and 15%+15% WSF treatment concentration levels of the individual hydrocarbons were generally lower and more variable in their ranked concentrations.

The individual hydrocarbons in each treatment group were categorized into their chemical classes and also ranked in order of decreasing rates of degradation (Tables 4b - 9b). A comparison of the hydrocarbon component concentrations to their ranked rates of degradation revealed similar results to those in the JET-A MFC, where the concentration of a hydrocarbon component did not determine its rate of degradation. However, in this microcosm experiment none of the hydrocarbon components that were ranked by concentration matched any of the hydrocarbons ranked by degradation rates in any of the

treatment groups (Tables 4b - 9b).

In the alkanes tetradecane was degraded at the fastest rate in the 5%, followed by the 15%+15%, the 15%, and the 0%+15% WSF treatments though it was the lowest or second to lowest in concentration. Tridecane was the second fastest alkane degraded, followed by decane and dodecane. The degradation rates for the higher molecular weight *n*-alkane hydrocarbons were faster than the rates of degradation for the lower molecular weight alkanes, similar to alkane degradation patterns in the JET-A MFC. The difference in this microcosm experiment was that the concentrations of the alkanes were much higher compared to the shorter chain alkanes.

A comparison of the ranked aromatic hydrocarbon concentrations to the ranked degradation rates showed more consistent results than in the JET-A MFC microcosm experiment. *m,p*-Xylene was highest in concentration in the 5%, 15%, 0%+15%, and the 15%+15% WSF treatments, but was consistently degraded at the second fastest rate in all treatment groups (Tables 6b and 7b). Similar to the JET-A aromatics, toluene was degraded at the fastest rate in all treatment groups and *o*-xylene was degraded at the slowest rate in the 15%, 0%+15%, and 15%+15% WSF treatments. At the 5% and 15%+15% WSF treatment levels the rank order of degradations were identical in both the JET-A and the JP-8 (Table 7a and 7b).

In the alkyl-substituted aromatics the concentration and degradation dynamics are very different from those in the JET-A MFC experiment. Propylbenzene was present in the highest concentration in the 5% and 15% WSF groups, but was degraded at the fastest rate in all treatment groups (Tables 8b and 9b). Butylbenzene was present in the highest concentration in the 0%+15% and 15%+15% WSF treatments, but was degraded at the slowest rate in the 5%, 0%+15%, and 15%+15% WSF treatment groups. Ethylbenzene was consistently the third highest in concentration, but was either the slowest in degradation as in the 15% WSF treatment, or was second to the slowest in the other three treatment groups (Table 9b). These results are inconsistent with the results in the JET-A MFC where ethylbenzene was degraded at the fastest rate in all treatment groups (Table 9a). Cyclooctane was present in the lowest concentration , but was degraded at the second fastest rate in the 5%, 0%+15%, and 15%+15% WSF treatments which was similar to its pattern of degradation in JET-A MFC.

Hydrocarbon Treatment Comparisons

In the alkanes tetradecane was degraded at the most rapid rate in the 5% WSF, followed by the 15%, the 15%+15%, and slowest in the 0%+15% WSF treatment (Appendix E). There was no significant differences in degradation rates between the

15% WSF and the two re-treated groups or between the two re-treated groups (Table 10) (Figure 22b). Tridecane was also degraded the fastest in the 5% WSF treatment, followed by the 15%+15%, 0%+15%, and 15% WSF treatments (Appendix E) (Figure 21b). Similar to tetradecane there were no significant differences in degradation rates between any of the treatment groups (Table 10).

Decane was degraded at the fastest rate in the 15%+15% WSF, followed by the 0%+15%, 15%, and 5% WSF treatment groups (Figure 19b) (Appendix E). In the 15% WSF treatment decane was degraded at a significantly different rate compared to both the 0%+15% and the 15%+15% WSF treatments. The 0%+15% and the 15%+15% WSF treatments were not significantly different from each other (Table 10).

Dodecane was degraded at fairly low rates in all treatment groups with the 15%+15% WSF being degraded the fastest, followed by the 0%+15%, 5%, and 15% WSF treatment (Appendix E) (Figure 20b). The degradation rates in the 15% WSF and the re-treated groups were all significantly different from each other (Table 10).

In the aromatic group of hydrocarbon components, toluene was consistently degraded at the fastest rate in all of the treatment groups, as in the JET-A MFC. The rate of degradation in the 15%+15% WSF was the fastest, followed by the 0%+15%, 5%, and 15% WSF treatments (Appendix E) (Figure 23b). The 15% WSF degradation rate was significantly different from the 0%+15% and the 15%+15% WSF treatments, but the 0%+15% and the 15%+15% WSF treatment groups were not significantly different (Table 11).

The mixture of *meta*- and *para*-xylenes were degraded in the identical order as toluene with the 15%+15% WSF being the fastest, followed by the 0%+15%, 5%, and 15% WSF treatments (Appendix E) (Figure 25b). There were significant differences in the degradation rates between the 15% and the 0%+15% and to the 15%+15% WSF treatments. The 0%+15% and the 15%+15% WSF treatments were also significantly different from each other (Table 11).

The degradation rates of benzene were 15%+15%, 0%+15%, 15%, and 5% WSF treatments (Appendix E) (Figure 24b). The 15% WSF treatment group was degraded at a significantly different rate than in the 0%+15% and the 15%+15% WSF groups, but not at different rates between the two re-treated groups (Table 11).

For *ortho*-xylene the patterns in treatment degradation rates were identical to those for toluene and the *meta*- and *para*-xylenes. The rank order in degradation rates for the treatment groups were 15%+15%, 0%+15%, 5%, and 15% WSF treatments (Appendix E) (Figure 26b). Similar to the other xylenes, there were significant

differences in the degradation rates between the 15% and the 0%+15% and 15%+15% WSF treatments, as well as between the 0%+15% and the 15%+15% WSF treatments (Table 11).

In the alkyl-substituted aromatic group propylbenzene was consistently degraded at the fastest rate in all of the treatment groups. The highest rate of degradation occurred in the 0%+15% WSF, followed by the 15%+15%, 5%, and 15% WSF treatments (Appendix E) (Figure 30b). There were significant differences between the 15% WSF and the 15%+15% WSF, but not between the 0%+15% WSF treatment. Both of the re-treated groups were not significantly different from each other (Table 12).

The degradation rates of cyclooctane were the 5% WSF, followed by the 15%+15%, the 0%+15%, and the 15% WSF treatments (Figure 28b) (Appendix E). There were no significant differences in the rates of degradation between the 15% WSF, the 0%+15%, and the 15%+15% WSF treatments (Table 12).

Butylbenzene was degraded at the fastest rate in the 15% WSF treatment, followed by the 0%+15%, the 15%+15%, and finally by the 5% WSF treatments (Figure 29b) (Appendix E). For butylbenzene, there was no significant differences between any of the three 15% WSF treatment groups (Table 12).

Ethylbenzene was degraded the fastest in the 15%+15% WSF, followed by the 0%+15%, 5%, and at the slowest rate in the 15% WSF treatment group (Figure 27) (Appendix E). All three treatment groups were significantly different from each other in their degradation rates (Table 12). A summary of the combined hydrocarbon components, ranked by their rates of degradation is listed (Table 14).

The metabolic hydrocarbon components that were produced during the course of the experiment also varied in concentration and were dependent on degradation of the parent compound and the rate of utilization of the metabolite by the microorganisms. The metabolites were propane, *cis*-2-pentene, pentane, 2-methylpropane, and 2,4-dimethylpentane (Figures 31b-35b, respectively). As in the JET-A MFC propane, pentane, 2,4-dimethylpentane, 3-methylpentane, *trans*-2-pentene and hexane displayed the same increases in concentrations at approximately the same time intervals of fifty, ninety-six, one hundred forty-four, and two hundred eighty hours. During the initial experiment more amounts of the compounds were released than from the later re-treated groups. For the hydrocarbon components 3-methylpentane, *trans*-2-pentene, and hexane there was a complete separation of the treatment responses between the initial treated groups and the re-treated microcosms (Figures 36b-38b).

JET-A MFC and JP-8 SAM Hydrocarbon Degradation Rate Comparisons

The accelerated degradation rates in the initial JET-A MFC water soluble fraction treatments, compared to the depressed degradation rates in the initial JP-8 water soluble treatments dominate the rankings of the hydrocarbon components within each microcosm experiment. In the JET-A MFC 0%+15% and the 15%+15% WSF treatments the degradation rates were significantly slower compared to the initial 15% WSF treatment. In the JP-8 SAM the re-treated microcosms generally had degradation rates significantly faster than their initial 15% WSF treatment degradation rates. There were sufficient patterns displayed in each microcosm experiment that indicate that microbial degradation mechanisms and metabolic pathways are similar when exposed to the same type of toxicant stressor.

An analysis of the initial microcosm treatment groups compared to each other does reveal some similarities between the degradation of hydrocarbon components in the two microcosms. In both microcosm experiments decane, benzene, and butylbenzene were all degraded at the most rapid rate in the 15% WSF, while dodecane, ethylbenzene, toluene, *meta*- and *para*-xylenes, and *ortho*-xylene were degraded more rapidly in the 5% WSF treatment (Table 15). The JET-A MFC rank order would indicate that the most rapid degradative rates occurred in the 15% WSF treatment. The JP-8 SAM would seem to indicate the reverse with the most rapid degradation occurring in the lower 5% WSF treatment.

A comparison of the MFC 0%+15% WSF degradation rates to the 15%+15% WSF rates indicated that only propylbenzene and butylbenzene were degraded faster in the 15%+15% WSF treatment than in the 0%+15% WSF treatment (Table 16). In the SAM experiment propylbenzene was the only hydrocarbon compound that was not degraded faster in the 15%+15% WSF treatment. An analysis of only the JET-A MFC results would indicate that in the 0%+15% WSF treatments the hydrocarbon components were degraded at the fastest rate. However, an analysis of the JP-8 SAM results would indicate the opposite relationship that the 15%+15% WSF treatment groups were degraded at faster rates.

Cyclooctane and tridecane were the only hydrocarbons that were not significantly different between the two microcosms in any of the treatment groups tested (Tables 17 and 19). Dodecane was the only hydrocarbon significantly different between the two microcosms in all treatment groups tested, with butylbenzene also being significantly different in all but two comparisons. The majority of the significant differences that were determined for the water soluble fraction treatments were between the 15% WSF treatments compared to each other in the MFC and the SAM microcosms and between the

Table 15. Initial treatment groups ranked in order of decreasing rates of degradation for each hydrocarbon component for the JET-A MFC and the JP-8 SAM experiments.

	JET-A MFC Rank			JP-8 SAM Rank	
Hydrocarbon	1	2	3	1	2
Decane	15%	1%	5%	15%	5%
Dodecane	5%	15%		5%	15%
Tridecane	15%			5%	15%
Tetradecane				5%	15%
Benzene	15%	5%	1%	15%	5%
Toluene	5%	15%	1%	5%	15%
m,p-Xylene	5%	15%	1%	5%	15%
o-Xylene	5%	15%	1%	5%	15%
Ethylbenzene	5%	15%	1%	5%	15%
Propylbenzene	15%	5%		5%	15%
Butylbenzene	15%	5%		15%	5%
Cyclooctane	15%	5%		5%	15%

Table 16. Re-treated initial treatment groups ranked in order of decreasing rates of degradation for each hydrocarbon component for the JET-A MFC and the JP-8 SAM experiments.

	JET-A MFC Rank		JP-8 SAM Rank	
Hydrocarbon	1	2	1	2
Decane	0%+15%	15%+15%	15%+15%	0%+15%
Dodecane	0%+15%	15%+15%	15%+15%	0%+15%
Tridecane	0%+15%	15%+15%	15%+15%	0%+15%
Tetradecane	0%+15%	15%+15%	15%+15%	0%+15%
Benzene	0%+15%	15%+15%	15%+15%	0%+15%
Toluene	0%+15%	15%+15%	15%+15%	0%+15%
m,p-Xylene	0%+15%	15%+15%	15%+15%	0%+15%
o-Xylene	0%+15%	15%+15%	15%+15%	0%+15%
Ethylbenzene	0%+15%	15%+15%	15%+15%	0%+15%
Propylbenzene	15%+15%	0%+15%	0%+15%	15%+15%
Butylbenzene	15%+15%	0%+15%	15%+15%	0%+15%
Cyclooctane	0%+15%	15%+15%	15%+15%	0%+15%

Table 17. Student's t test for significant differences ($*t_{0.05\,(2),\,26}$) between log regressed degradation rate coefficients for each alkane hydrocarbon component when compared between the two microcosm experiments in the 15% WSF, the 0%+15% WSF, and the 15%+15% WSF treatment groups.

	JP-8 SAM		
JET-A MFC (WSF)	15% WSF	0%+15% WSF	15%+15% WSF
Decane [15%]	*	*	*
Decane [0% + 15%]	*		
Decane [15% + 15%]	*		
Dodecane [15%]	*	*	
Dodecane [0% + 15%]	*	*	*
Dodecane [15% + 15%]	*	*	*
Tridecane [15%]			
Tridecane [0% + 15%]			
Tridecane [15% + 15%]			
Tetradecane [15%]			
Tetradecane [0% + 15%]		*	
Tetradecane [15% + 15%]			

Table 18. Student's t test for significant differences ($*t_{0.05\,(2),\,26}$) between log regressed degradation rate coefficients for each aromatic hydrocarbon component when compared between the two microcosm experiments in the 15% WSF, the 0%+15% WSF, and the 15%+15% WSF treatment groups.

	JP-8 SAM		
JET-A MFC (WSF)	15% WSF	0%+15% WSF	15%+15% WSF
Benzene [15%]	*	*	*
Benzene [0% + 15%]	*		
Benzene [15% + 15%]	*	*	*
Toluene [15%]	*		
Toluene [0% + 15%]	*		
Toluene [15% + 15%]	*	*	
m,p-xylene [15%]	*		
m,p-xylene [0% + 15%]	*		*
m,p-xylene [15% + 15%]	*		*
o-xylene [15%]	*	*	*
o-xylene [0% + 15%]	*	*	*
o-xylene [15% + 15%]	*		

Table 19. Student's t test for significant differences ($*t_{0.05\,(2),\,26}$) between log regressed degradation rate coefficients for each alkyl-substituted aromatic hydrocarbon component when compared between the two microcosm experiments in the 15% WSF, the 0%+15% WSF, and the 15%+15% WSF treatment groups.

	JP-8 SAM		
JET-A MFC (WSF)	15% WSF	0%+15% WSF	15%+15% WSF
Butylbenzene [15%]		*	*
Butylbenzene [0% + 15%]		*	*
Butylbenzene [15% + 15%]	*	*	*
Cyclooctane [15%]			
Cyclooctane [0% + 15%]			
Cyclooctane [15% + 15%]			
Ethylbenzene [15%]	*	*	*
Ethylbenzene [0% + 15%]	*		
Ethylbenzene [15% + 15%]			
Propylbenzene [15%]			
Propylbenzene [0% + 15%]			*
Propylbenzene [15% + 15%]			*

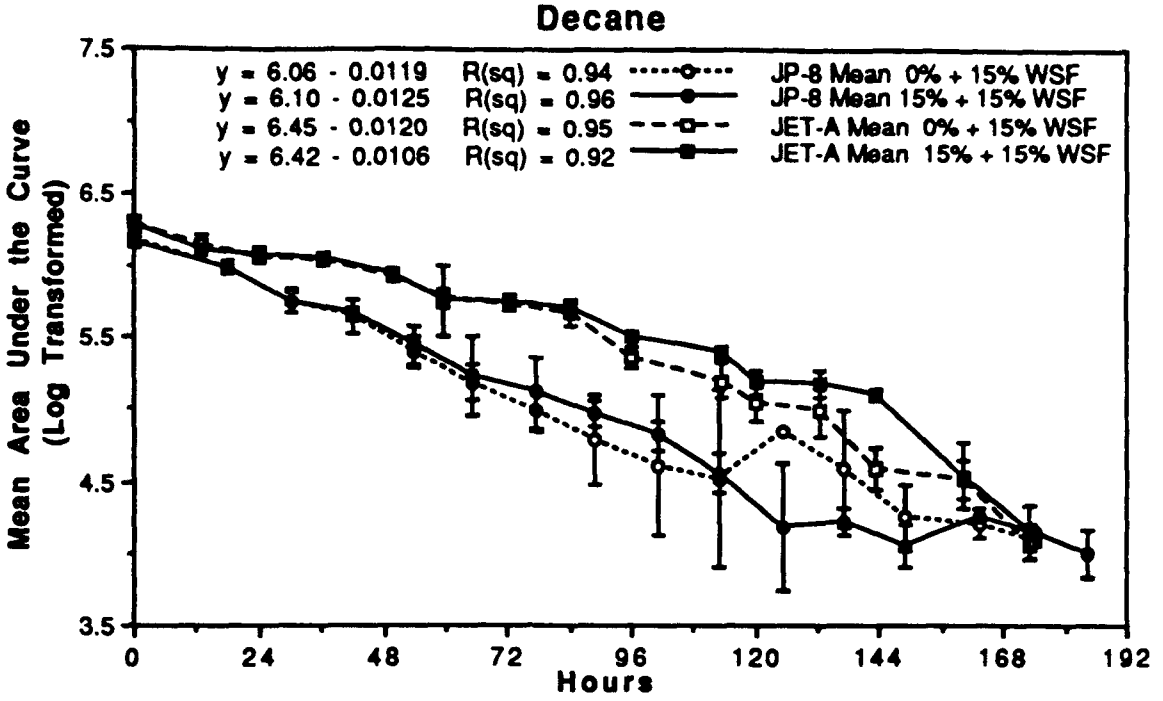


Figure 39. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for decane in the re-treated microcosm groups.

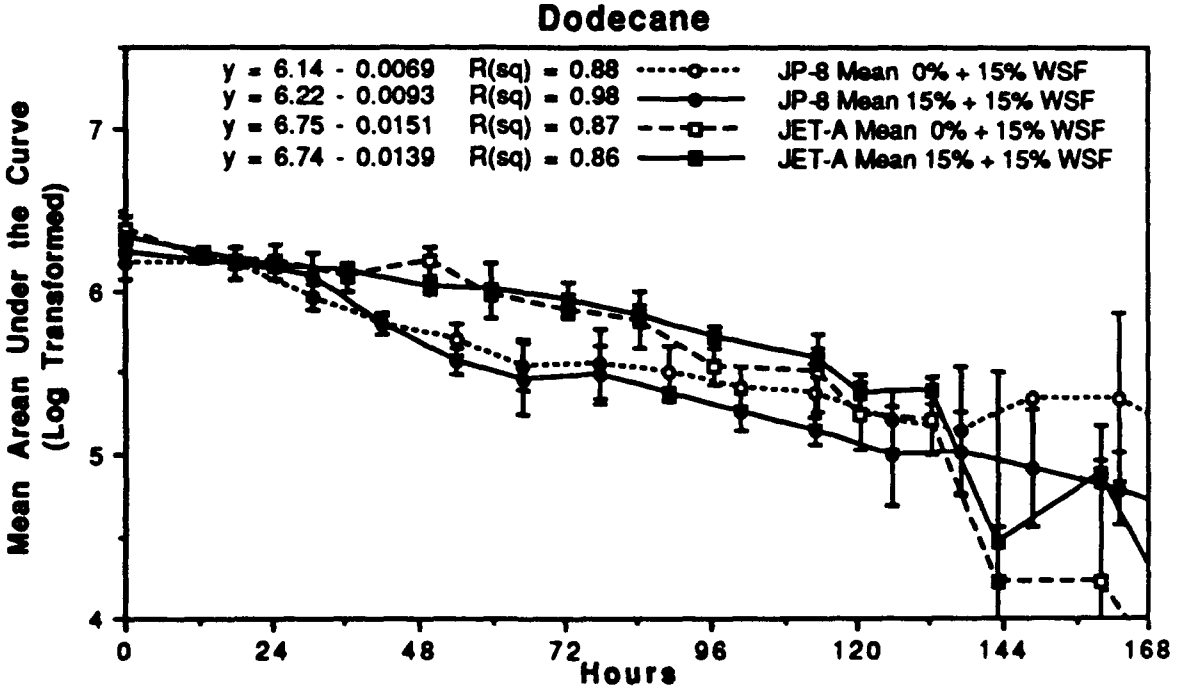


Figure 40. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for dodecane in the re-treated microcosm groups.

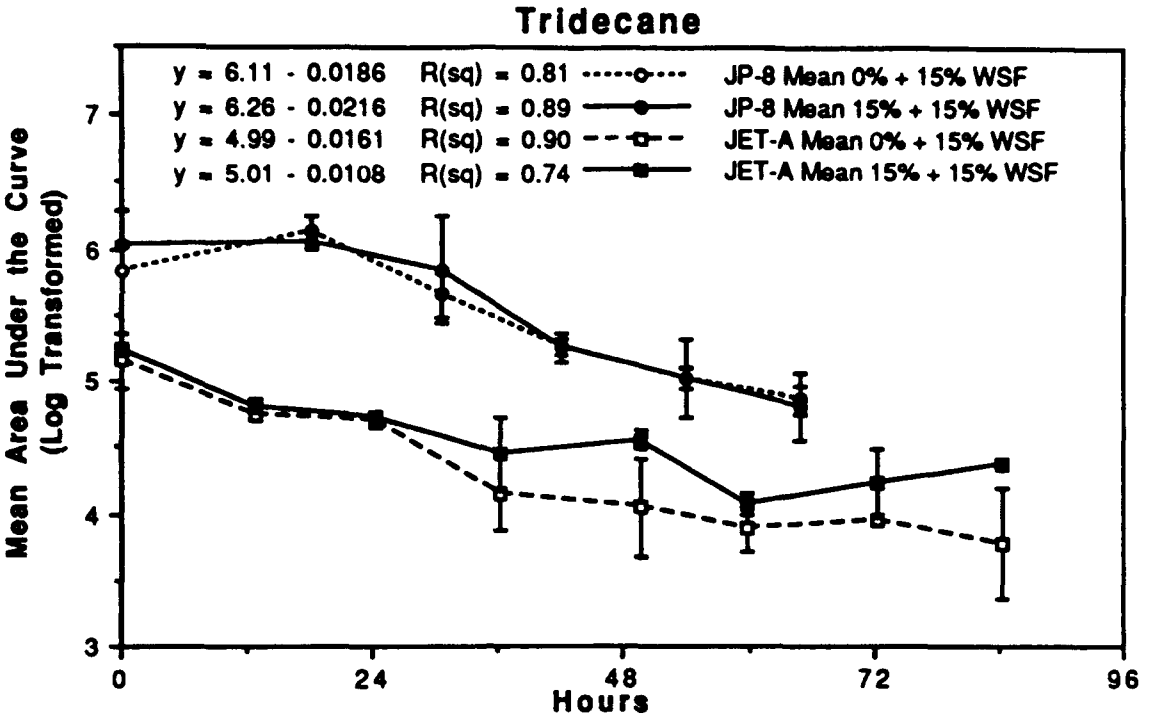


Figure 41. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for tridecane in the re-treated microcosm groups.

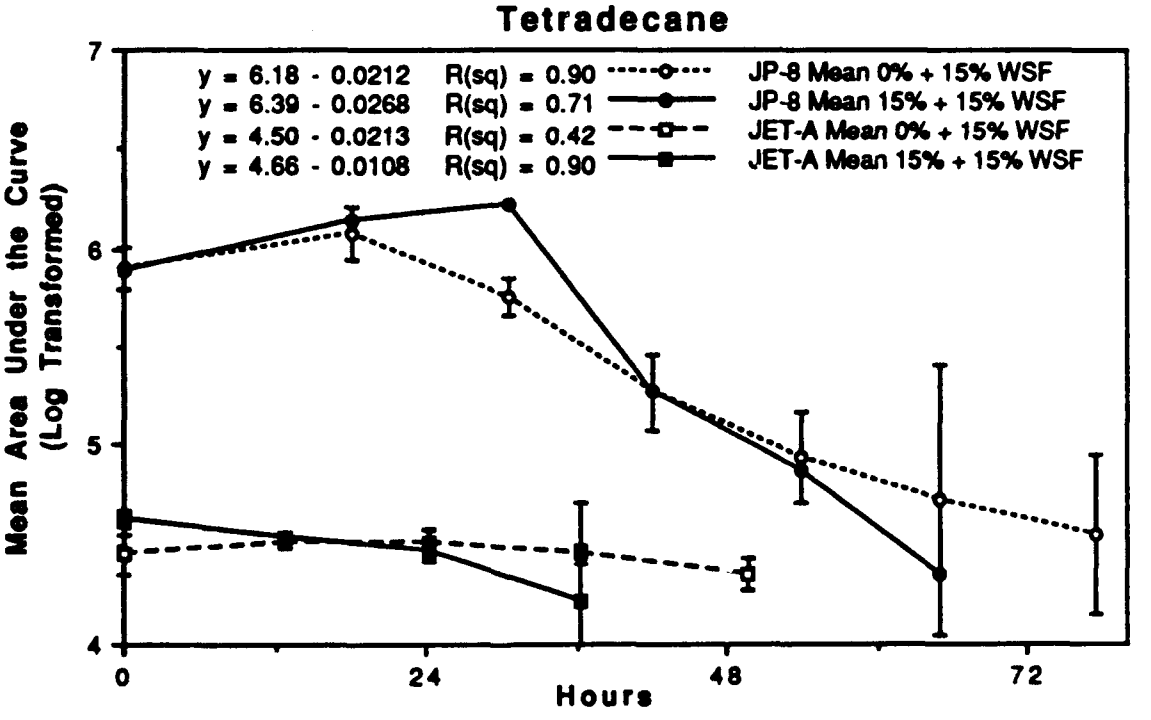


Figure 42. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for tetradecane in the re-treated microcosm groups.

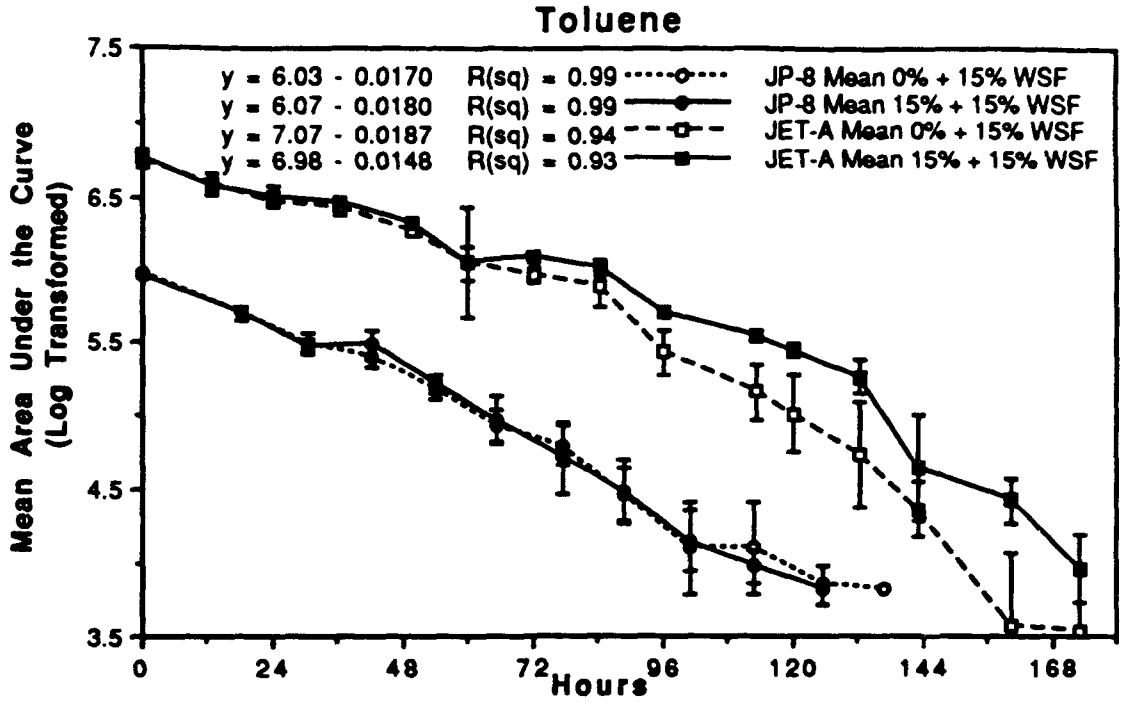


Figure 43. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for toluene in the re-treated microcosm groups.

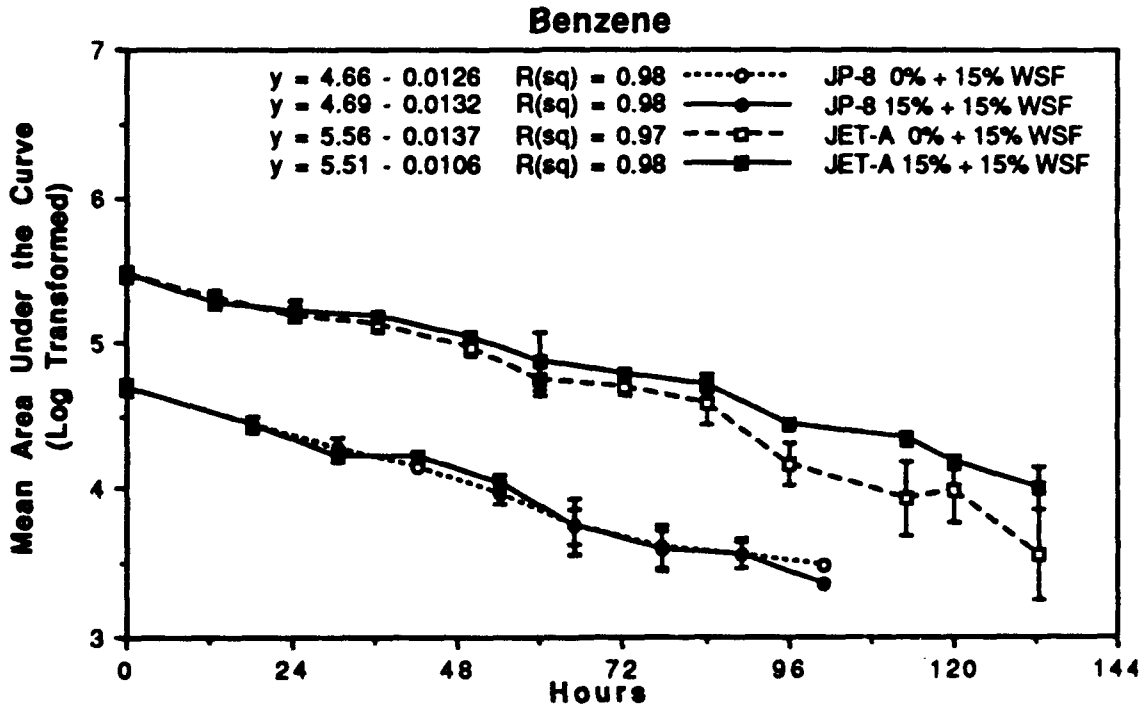


Figure 44. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for benzene in the re-treated microcosm groups.

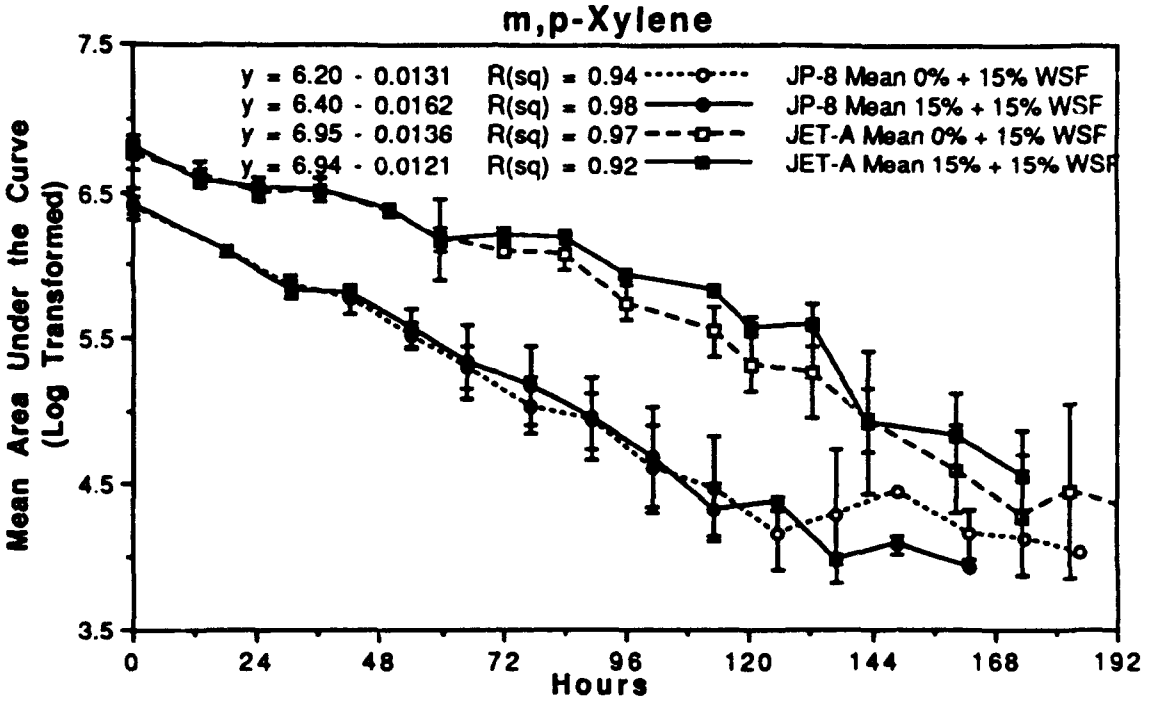


Figure 45. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for *m,p*-xylene in the re-treated microcosm groups.

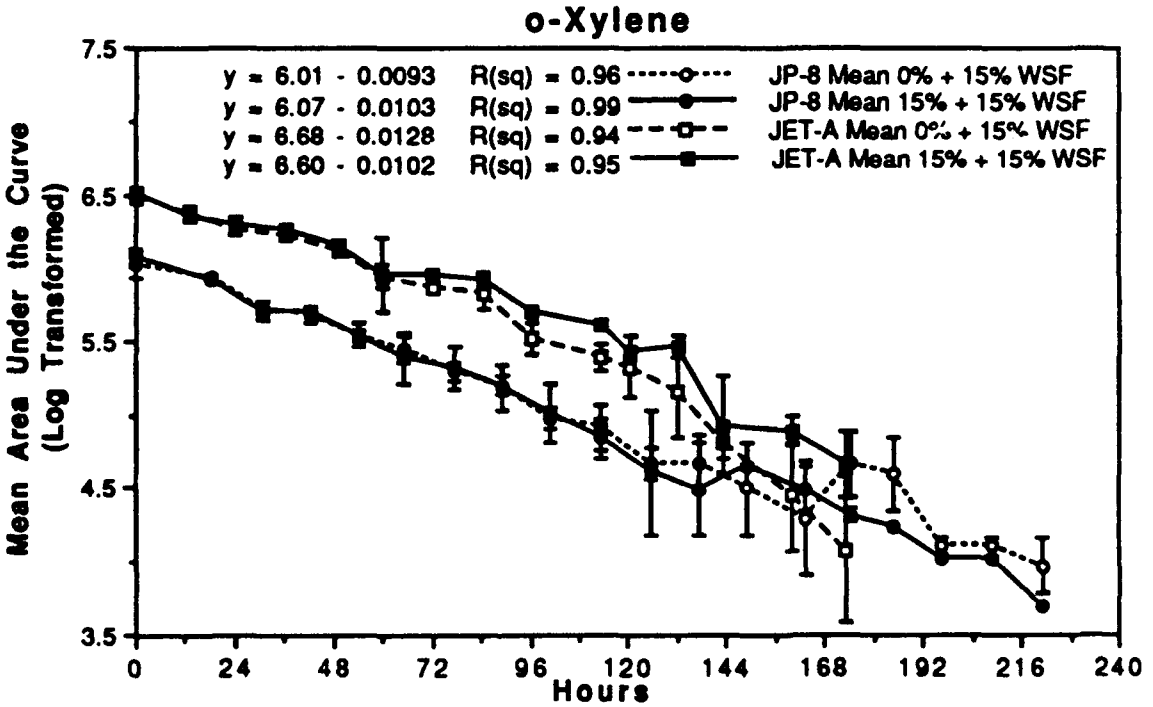


Figure 46. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for *o*-xylene in the re-treated microcosm groups.

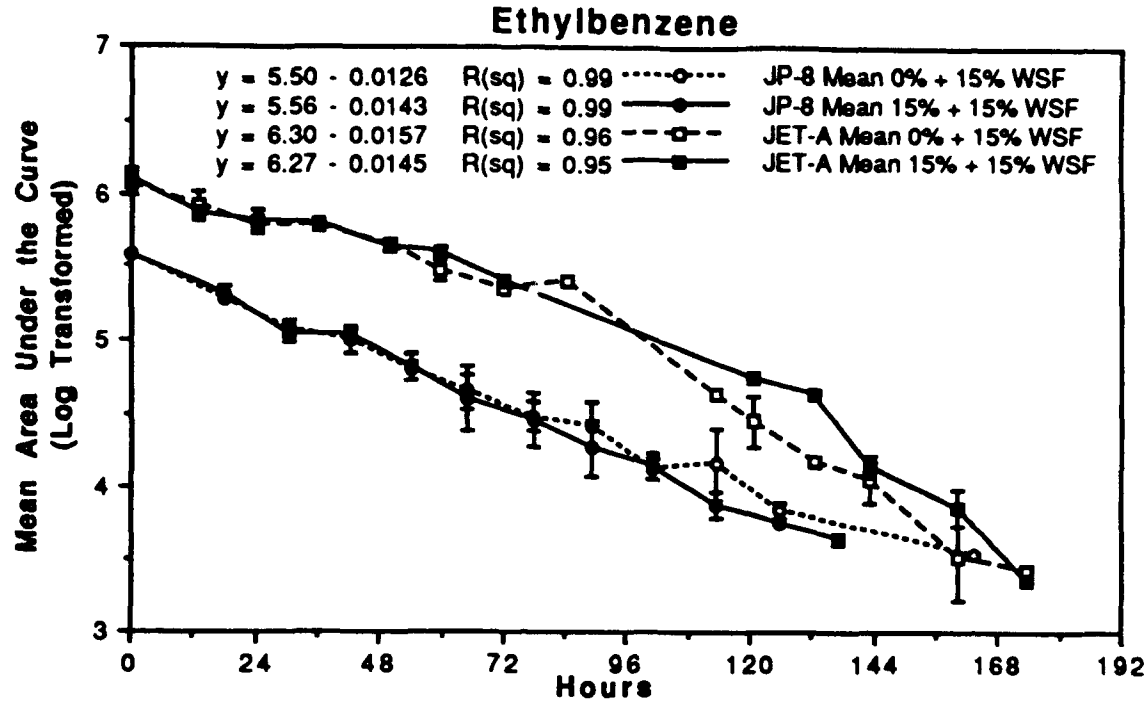


Figure 47. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for ethylbenzene in the re-treated microcosm groups.

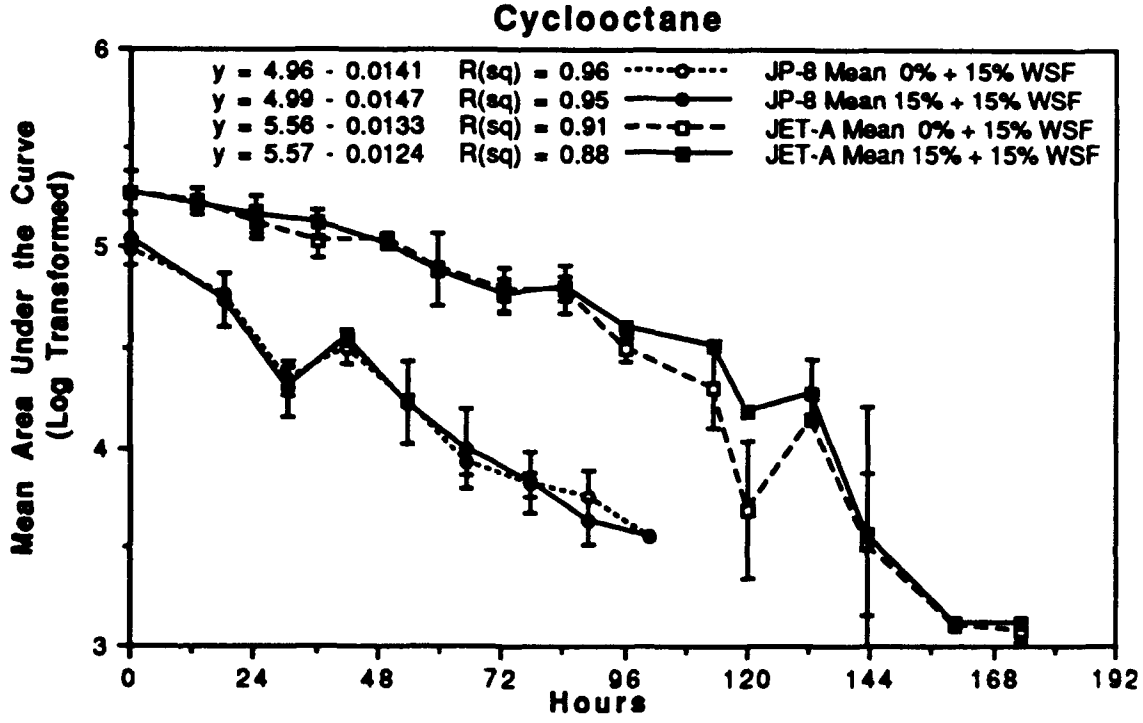


Figure 48. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for cyclooctane in the re-treated microcosm groups.

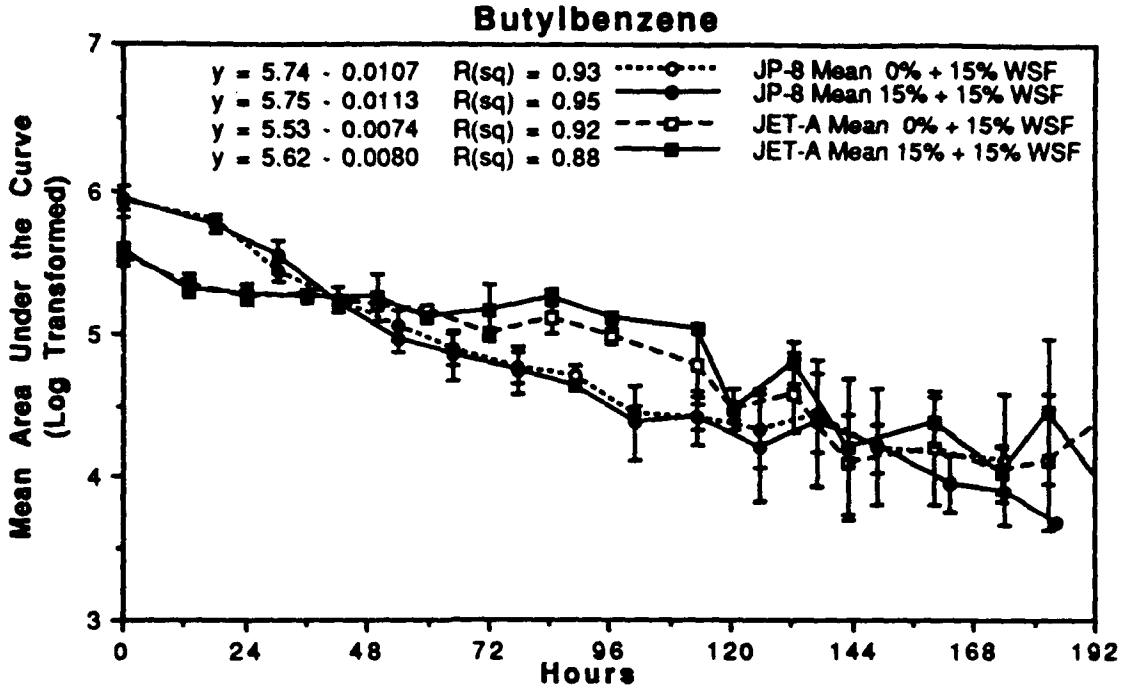


Figure 49. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for butylbenzene in the re-treated microcosm groups.

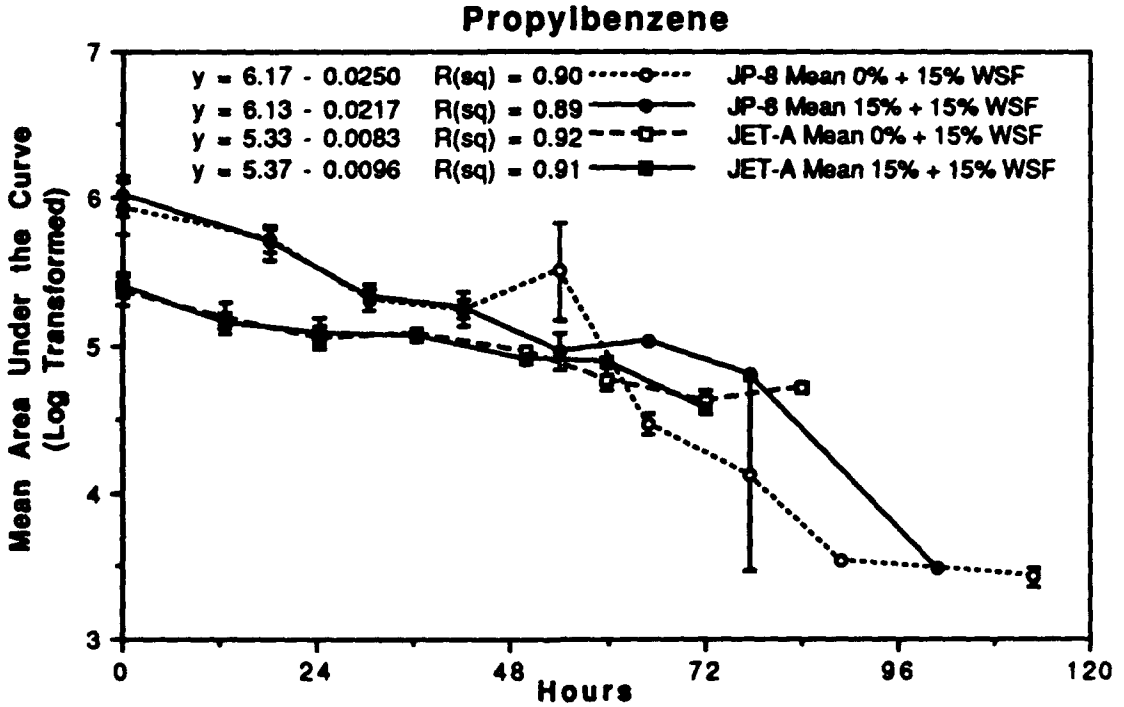


Figure 50. Comparison of JET-A MFC and JP-8 SAM mean degradation rates for propylbenzene in the re-treated microcosm groups.

re-treated 0%+15% WSF and the 15%+15% WSF treatments in the two microcosms (Tables 16-19).

Discussion

The Mixed Flask Culture microcosm and the Standardized Aquatic Microcosm are both used to model ecosystem structural and functional processes for determining the effects of chemical toxicants on aquatic environments. The validation of these microcosms for use in ecological risk assessments has been a slow process and is still subject to a degree of uncertainty. The principal concerns involve which properties to measure that "best" constitute ecosystem processes, the degree of structural realism that should be replicated, the interpretation of the multivariate data results, the replicability of the test systems, and the applicability of the results for extrapolation to the ecosystem. The decisions that are made to address these issues will dictate the accuracy and validity of risk assessments to evaluate and predict the effects of toxicants on an ecosystem from extrapolated microcosm results.

The decision to use one type of microcosm model for ecosystem-level testing, as opposed to another has previously been dependent on the investigator's opinion of what is an ecosystem and what properties or functions define that ecosystem. The microcosm models developed and used have been evaluated by the researchers conducting the tests, but few comparative evaluations have been made between microcosm experiments. Fewer studies have been conducted to compare the effect responses between different microcosm models using the same type of toxicant. The MFC and the SAM were selected for comparison due to their generic similarities, but distinctly different theoretical and structural histories. Their specific attributes make them ideal systems to compare and determine whether effect intensities measured in one type of environment are similar to other ecosystems.

Intuitively, the use of real assemblages of organisms excised from a system and treated with a chemical contaminant is an obvious and rational procedure to use for examining community and potential ecosystem-level effects. The principal difficulty with this kind of approach is attaining the degree of realism that is economically and experimentally feasible. The cost associated with the construction, maintenance, operation, monitoring, and staffing required to conduct a "real-world" microcosm experiment is prohibitive to most researchers. Every decision that is made to modify the system to be simpler, cost effective, and less labor intensive will compromise the realism of the system. The dilemma becomes how much realism can be sacrificed without jeopardizing the integrity of the experiment. In conjunction with these decisions is the problem of how to replicate these systems to confirm test results and to reproduce these results in other geographical areas that have different indigenous

assemblages of species.

The use of a synthetically constructed microcosm has the distinct advantage of not even addressing the "realism" dilemma. The parameters of interest focus on the functional properties, effect responses, and the interactions of the organisms through time within the contained system. The microcosms attempt to simulate only the general properties and functions present in all natural systems, not the actual structural components. There has been an inherent bias against using artificially contrived microcosms, however, these systems have proved to demonstrate many of the same functional processes and dynamics observed in natural environments. In addition, the replicability and reproducibility of these systems increase their versatility for use in other types of studies.

The final decision to select one microcosm type as opposed to another should depend on the specific hypotheses to be tested and the appropriateness of the microcosm type for testing the hypothesis. The selected microcosm type should be able to display the interrelationships and rate responses that have been observed in the ecosystem being investigated. The selection of a microcosm model that is functionally dependent on photosynthesis/respiration processes may not be the appropriate system to use if the hypothesis to be tested relates to detrital community processes and structure. Extrapolations from responses obtained using inappropriate microcosm tests to predict effects at the ecosystem-level of organization will be limited in applicability and in many cases inaccurate.

The MFC and the SAM are compared to test whether the two systems display similar degradative rate responses, with the same level of intensity, when treated with a similar type of toxicant. As valid generic models of ecosystem properties and dynamics they are expected to display similar patterns in the rates of degradation when abiotic environmental conditions are maintained at similar and constant levels. The microbial communities in each system are assumed to be composed of similar types of microorganisms that perform similar types of metabolic and degradative processes. The specificity of microbially-mediated enzymatic degradative pathways will dictate that the individual components in the toxicant will be degraded at rates specific for the chemical structure and properties of that component. Compounds with similar chemical structures and properties would be expected to be degraded by similar enzymatic pathways at relatively similar rates.

The actual rates of degradation for each component may be different when compared between the two microcosms, due to the variability of utilization rates inherent in all populations, but they should not be significantly different. The

populations are distinctly different and derived from distinctly different sources, but their rate of component utilization and degradation (slopes) through time should display similar rate patterns. The similarity of the constraints placed on the two microcosm systems in terms of environmental conditions and the type of toxicant should elicit similar patterns of responses, regardless of the actual species composition. The existence of universal ecosystem properties and universal patterns of response would imply that these rate patterns determined in the microcosms should be similar to patterns observed in natural ecosystems.

The comparison of the MFC and the SAM degradation rate patterns to field observations is more difficult due to the variability of previous sampling methodologies and the stochastic environmental conditions that dramatically affect the microbial rates of degradation. The consistent pattern that has been observed and documented in both field and laboratory microbial degradative studies is that microbial communities pre-exposed to hydrocarbon mixtures will degrade subsequent hydrocarbon mixtures at quantifiably faster rates (Aelion et al., 1989; Atlas, 1981; Evans, 1991; Focht, 1988). A re-treatment of the MFC and the SAM should cause the degradative rate responses (slopes) to display similar patterns of increased degradation rates. This criteria is used to determine whether the MFC or the SAM display generic functional processes that are comparable to field determined results.

The validation of the MFC or the SAM to accurately simulate ecosystem functional rate responses and patterns would establish their use in ecological risk estimates and chemical hazard assessments. The issue of whether microcosms must resemble real ecosystems as closely as possible to be valid models of ecosystem dynamics would be resolved. Ecological realism and complexity in microcosms may not be necessary to discern and reveal ecosystem-level functional processes. Specific populations would not determine general ecosystem properties or rate responses to stress and the implications would be that there may not exist a specific organism or factor that can be used to indicate ecosystem responses to toxicant stress.

The Mixed Flask Culture Microcosm

The mixed flask culture microcosm design attempts to incorporate some degree of realism by using "real ecosystem" assemblages of organisms. These natural assemblages are presumed to reflect more "natural" responses to toxicant exposure than a synthetically assembled microcosm system. The primary limitation of these microcosms is their smaller size which places added constraints on the full expression of functional and structural responses.

In the construction of the MFC's, the stock community is inoculated into relatively sterile 1 L test chambers containing autoclaved sand and sterile T82MV medium. At this stage the microcosms display early successional colonization dynamics. Initially, there are few species interactions, the cycling of nutrients and energy flows are not complete, organism numbers and species diversity are still very low, and the microbial communities are relatively inactive. During the six week equilibration period, the microcosms are re-inoculated with fresh aliquots from the stock aquarium, cross inoculated between the microcosms, and evaporative losses replenished with fresh medium. In essence, the successional phase is manipulated to produce an established and productive community of organisms, similar in organizational structure and complexity to the stock community.

At the time of toxicant addition, the MFC's have been manipulated to be extremely productive, high in available nutrients, and with very high cycling rates for carbon and other energy-rich compounds. The biomass and total number of organisms present are artificially high, species diversity is very high, and there is a very active microbial degradative and decompositional community in place. The detrital matter is composed of high molecular weight organic material that is high in nutritive quality. Algal cells and cell fragments, molts of cladocera, amphipods, and copepods, fecal material containing intact and fragmented cellular material, and flocculent clumps of colloidal cellular and fecal debris held together in gelatinous matrices formed by bacterial activity are present. The silica sand sediment is discolored and covered with greenish-yellow biofilms and organic coatings to form loosely consolidate aggregates. Rapid physical and chemical degradation and decomposition of organics is possibly due to complex and diverse microbial populations.

The initial conditions in the MFC's have been controlled to such an extent that their nutritive status, diversity of organism activity, and carrying capacities are well beyond the "real world" levels for a system of that size. In any natural system there are always species that are dormant, inactive, or controlled to some extent by competition and predation. They all have the potential to become viable and contribute to the structure and dynamics of the system during optimal conditions. In the MFC microcosms, the luxury supply of nutrients and the temperate, environmentally controlled conditions have allowed many of the otherwise dormant species to become active and to temporarily compete for nutritive and spatial resources.

The addition of a toxicant to these highly optimized systems elicits responses that are not necessarily applicable to the "real" ecosystem responses that the microcosms were intended to simulate. The hydrocarbon components in the water soluble fractions of

the jet fuel provided additional carbon sources and alternative metabolic pathways for utilization by greater consortia of microorganisms. Degradative rates for the hydrocarbon components were relatively high and involved active mineralization processes as well as cometabolic processes by highly developed, interactive microbial consortia (Figures 19-30). In addition, the detection of hydrocarbon intermediates prior to treatment in the MFC microcosms indicate that the algae were producing hydrocarbon intermediates and aromatic compounds that the microbial population were able to degrade. The rapid degradation of the water soluble components in the MFC microcosms could have also been a function of presence of microorganisms pre-adapted to utilize and degrade the similarly structured aromatic compounds. The initial conditions in the MFC microcosms will alter the degradation rates of the hydrocarbons and is apparent when compared to the degradation rates in the re-treated MFC microcosms that were not amended with fresh medium, cross-inoculated, or re-inoculated prior to the second treatment.

In the "mature" (re-treated) MFC microcosms there were more total numbers of individuals, but they represented fewer species of organisms. The algal species were dominated by the blue-green algae and several species of *Scenedesmus* that formed small, dark green clumps in which other bacterial organisms, rotifers, amoebae and protozoa were associated. The detritus was composed of low molecular weight organic material that was highly fragmented and low in nutritive value. The debris contained less intact particulates and consisted primarily of denser, less flocculent clumps of unidentifiable cellular and fecal debris that were more yellowish-brown in color compared to the vibrant yellowish-green color in the initial microcosms. The ostracod detritivore populations were also very high. The rates of degradation of the hydrocarbon components in these "aged" systems were slower than the initial microcosm experiment treatment groups with the slowest rates consistently occurring in the 15%+15% WSF treatments (Table 16).

The Standardized Aquatic Microcosm

The SAM is an artificially constructed system with the chemical, physical and biological components assembled to meet specifically defined and exact criteria. The organisms were selected due to their availability and diversity of metabolic pathways that are representative of generic functional groups (Taub, 1984). The SAM's were developed to be reproducible, non-site specific assemblages that demonstrate ecosystem structural and functional properties, rather than ecosystem structural reality. Biological relationships and responses are believed to be more apparent in these

simplified systems without the confusion associated with multispecies assemblages that may otherwise hide important processes and dynamics (Taub, 1984).

The SAM microcosms are sterile at the initiation of the test with the inoculation of the algae occurring several days prior to the inoculation of the protozoa and small macroinvertebrates. However, the addition of the toxicant occurs within one week of the construction of these systems that were still in the early successional stages of development. The short time frame between the construction, inoculation, and treatment of the SAM's causes these systems to be relatively devoid of both prokaryotic and eukaryotic organisms, with no carbon reservoir at the initiation of the test. The large nutrient source is limited to the chemically defined liquid medium (T82MV) and is only available to the algae for utilization. The bacteria inoculated into the medium with the protozoa are chemoorganotrophs and incapable of utilizing the inorganic nutrient salts in the medium.

The organic detrital matter consisted of the 0.5 g of cellulose and chitin that were added to the microcosm sediment on the day of construction and any algal cells, cellular debris, fecal matter, and molts that had accumulated in the first seven days. The limited quantity and quality of this initial organic detrital matter was capable of supporting the initial low densities of bacteria, protozoa, rotifers, and ostracods. The long-term effects of the initial poor quality of the detrital matter and lack of an available carbon reservoir on the extended success of these populations is still unresolved. It was observed that not until four to six weeks had elapsed into the experiment that detritivore populations were present to a significant degree (Landis et al., 1993, 1994). The only microorganism present at the initiation of the experiment was the bacteria *Enterobacter aerogenes*. This organism is used as the food source for the laboratory cultures of rotifers and was, by its association to the organisms, inoculated into the microcosms at the same time. Airborne microorganisms eventually entered the microcosms and supplemented the bacterial community. In addition, the overall scarcity of organisms in relation to algal biomass created an initial system with little or no interaction between the species present. The initial rate of detrital utilization and cycling of energy and carbon were still in the preliminary stages and the biochemical metabolic pathways for transformation and degradation processes were limited or non-existent. The lag time in the degradation of the hydrocarbon aromatics may have also been due to the lack of a pre-adapted microbial community to biogenically produced hydrocarbons or the lack of a diverse microbial community existing in these microcosms that were both present in the initial MFC microcosms.

The responses elicited from the exposure of the toxicant in the initial SAM

systems are presumed to be highly sensitive (Kindig et al., 1982). The relatively "young age" of the systems implies that they have not developed overlapping functional and structural components. They are presumed to be unable to adapt to the stress event and use alternative pathways to dampen the potential measurable effects of the toxicant. An alternative viewpoint is that the initial SAM's are so disconnected in organism interactions and cycling processes that they simulate individual single species toxicity tests that happen to be conducted in one container at the same time. The initial direct effects of the toxicant on the few organisms present are expected to yield simple mortality results that lack ecological meaning. The indirect effects on subsequent generations of the organisms, the quality and quantity of organic detrital matter produced, and the cycling of carbon, energy, and nutrients will be more subtle and difficult to extrapolate at the ecosystem-level. The initial SAM conditions are so artificial compared to natural ecosystems that only an extreme catastrophic event will create a similar system as sterile and barren of an organic carbon reservoir.

The initial conditions in the SAM's resulted in slower degradation rates of the water soluble fractions of hydrocarbons compared to the re-treated SAM microcosms (Figures 19b-30b). The initial conditions in the "aged" SAM's prior to re-treatment were very different from the initial SAM's. They were characterized by trophic dynamics that were highly interactive and included competition for limited food resources and predation. Daphnia populations were very high and the remaining algae were the blue-green algae *Lyngbya* sp. and *Anabaena cylindrica* and the green alga *Scenedesmus obliquus.* Ostracods, rotifers, and microbial communities were more developed, interactive, higher in abundance, and actively processing the organic matter that had accumulated during the course of the experiment. The detrital matter consisted of yellowish-brown, low molecular weight organic matter consisting of unidentifiable re-cycled cell fragments, molt fragments and fecal material. The silica sand sediment was discolored yellowish-green and slightly coated with organic films. The degradation rates of the hydrocarbon components in these microcosms were significantly faster when compared to their counterparts in the initial treatments, with the 15%+15% WSF treated components consistently being degraded at faster rates (Table 16). The increased rate of hydrocarbon degradation in the re-treated SAM's compared to the initial degradation rates agrees with the criteria selected to establish the validity of the SAM to display ecosystem-level properties.

JET-A and JP-8 Hydrocarbon Comparisons

The hydrocarbon degradation rates in the initial MFC and the SAM WSF treatment

groups are significantly different from each other. The differences appear to be more dependent on the initial conditions that existed in each of the microcosms, rather than on the concentrations or compositions of the water soluble fractions they were treated with (Tables 2-9, 13, and 14). In both the JET-A and the JP-8 the WSF's of jet fuel are composed primarily of the same hydrocarbon components. The differences are in the concentrations of those components and their ranking in concentration as a specific chemical class of hydrocarbons in the fuel mixture. The concentrations of the alkane, aromatic, or alkyl-aromatic hydrocarbon class of compounds will determine the types of microbial populations degrading the components, not the concentrations of the individual hydrocarbon components (Walker et al., 1968; Westlake et al., 1974). The rates at which the microbial populations degrade these components becomes dependent on the initial functional and structural conditions within a system, the chemical structure and properties of the hydrocarbons, and lastly on the concentration of the individual components.

The hydrocarbon components in the highest concentration in the JET-A fuel were the mono-aromatics, followed by the alkyl-substituted aromatics. The alkanes were variable in concentrations with dodecane in the highest concentration, of all the compounds and tetradecane in the lowest (Table 2). The hydrocarbons that were degraded at the fastest rates were the aromatic compounds, followed by the alkanes, and last by the alkyl-aromatics (Table 13). Though there was some overlap between the classes of hydrocarbon degradation rates, as a group they were degraded at distinctly different rates.

The majority of the active hydrocarbon degrading microbial populations in the JET-A MFC microcosms were specific for aromatic oxidation, cooxidation, and cleavage of the aromatic ring structure (Atlas, 1981; Atlas and Bartha, 1994; Focht and Westlake, 1988; Gibson, 1974, 1977). The higher concentration of the aromatic in the JET-A water soluble fraction mixture selected for these organisms. The microorganisms capable of degrading alkanes were also present, but their activity was secondary to the aromatic degrading microorganisms, due to the lower concentrations of available alkane substrates for growth and reproduction (Brock et al., 1994). The slowness of the degradation rates for the alkyl-aromatic compounds compared to the degradation of the *n*-alkanes is due to their greater structural complexity that inhibits the initial microbially induced oxidative attack, subsequent oxidation, and cleavage of the aromatic ring.

In the JP-8 fuel mixture the alkanes with ten to fourteen carbon atoms in length were present in the highest concentrations, with the alkyl-substituted aromatics next,

and the mono-aromatics in the lowest concentration (Table 3). The rates of degradation were more rapid for the longer chain alkanes tetradecane and tridecane, with decane and dodecane being degraded at the slower rate in almost all of the treatment groups. The alkyl-substituted aromatics were degraded at slightly slower rates and the aromatics at the slowest rates (Table 14). The higher concentration of alkanes in the JP-8 WSF supported microbial populations that were primarily more generalized, opportunistic species. These organisms require fewer specialized enzymatic mechanisms to degrade the simple, straight chain carbon structures (Atlas and Bartha, 1994; Pirnik et al., 1974; Walker et al., 1976b). The presence of these organisms also enabled the degradation of the alkyl-substituted chains on the aromatic compounds to occur at faster rates. The microorganisms that are capable of utilizing and degrading the aromatic compounds were present, but due to the lower concentration and importance of the aromatics in the water soluble fraction, their growth and reproduction were reduced in comparison (Brock et al., 1994).

Hydrocarbon Degradation Similarities

The degradation of aromatics requires specialized biochemical mechanisms to cleave the aromatic ring structure and these specialized mechanisms are only provided by specialized microbial communities (Atlas and Bartha, 1993) (Appendix A). This requirement for specialized enzymatic mechanisms will determine the rate at which these transformation and degradative processes occur. In the MFC these specialized organisms would be present and active at the time of treatment, due to the production of biogenic hydrocarbons by the algae. In the SAM these organisms would have had to enter the microcosms via the algae and other laboratory culture inocula, or transported into the microcosms adsorbed on the surfaces of airborne dust and soot particles. The role of the photoautotrophic bacteria (the blue-green algae) already in place in both microcosms was not investigated, but may have also assisted in the degradation of the aromatic compounds.

The microorganisms capable of degrading the aromatics in the MFC and the SAM would be expected to perform the chemical transformation at rates that should be relatively similar, due to the specificity of the degradative pathways. A comparison of the aromatic degradation rates in the two microcosms did display similar rate patterns. Toluene was always degraded at the fastest rate in all treatment groups, benzene was degraded at the slowest rate in both the MFC and SAM 5% WSF treatment groups, and o-xylene was always degraded at the slowest rate in all of the 15% WSF treatment groups as well as in the re-treated groups (Tables 7, 13-14). When the rates of degradation in

each treatment group were compared benzene and ethylbenzene were degraded at the fastest rate in both of the 15% WSF treatment microcosms and slowest in the 5% WSF groups. Conversely, toluene, *m,p*-xylene, and *o*-xylene were degraded at the fastest rate in both microcosms in the 5% WSF group and slowest in the 15% WSF treatment groups (Table 13).

The Student's t test was used to compare the individual hydrocarbon's degradation slopes both within each microcosm experiment and between the two microcosms. Within each microcosm experiment the hydrocarbon degradation rates in the re-treated 0%+15% and the 15%+15% WSF treatment groups were more similar to each other than they were to the initial WSF treatment groups. The magnitude of the significant differences were principally between the initial 15% WSF treatment groups compared to the 0%+15% and the 15%+15% WSF treatment groups. The significant differences between the two re-treated groups were to a much less extent (Tables 10-12).

A comparison of the hydrocarbon degradation rates between the MFC and the SAM indicate that the SAM 15% WSF treatment group accounts for the majority of the significant differences. The degradation rates of benzene, toluene, *m,p*-xylene, *o*-xylene, and ethylbenzene in the SAM 15% WSF treatment group were significantly different from the SAM re-treated microcosm groups and the MFC 15%, 0%+15% and 15%+15% WSF treatment groups (Tables 11-12 and 18-19). When the two 0%+15% WSF treatment groups were compared using the same compounds, they were not significantly different except for *o*-xylene. In the 15%+15% WSF treatment groups only benzene was significantly different (Table 18).

The degradation rates in the initial SAM 15% WSF treatments were very different from both the re-treated SAM's and all of the MFC WSF treatment groups. If the re-treated microcosms had also displayed significantly different degradation rates between the MFC and the SAM, then the results may indicate that the rates were a function of the microcosm type and the jet fuel composition. However, the similiarity of the 0%+15% WSF treatments and the 15%+15% WSF treatments between the two microcosms indicate that the degradation of these compounds follow similar metabolic pathways and that the two microcosms were more similar to each other in functional responses, after an initial conditioning period had elapsed. The lack of significant differences between the MFC 15% WSF treatments compared to the SAM 0%+15% and 15%+15% WSF treatments implies that the functional processes in the initial MFC's were also more similar to the functional processes occurring in the re-treated SAM microcosms.

The alkanes and the alkyl-substituted aromatics in the re-treated groups also

displayed similar degradation rates when the MFC and the SAM 0%+15% and the 15%+15% WSF were compared (Tables 17 and 19). Dodecane, tetradecane, and butylbenzene were the only hydrocarbons significantly different in the 0%+15% WSF treatments, while in the 15%+15% WSF treatment groups the only components significantly different were dodecane, butylbenzene, and propylbenzene. At these concentration levels, the implications are that the microorganisms responsible for the degradation of specific hydrocarbon classes of compounds will utilize principal enzymatic systems and specific degradative pathways. The relative consistency and similarity of the degradation rates for each class of hydrocarbons, regardless of the actual composition of the microbial community or the composition of the jet fuel support these results.

The metabolic by-products or intermediates formed during the degradative processes in all treatment groups in both microcosms were also very similar. These compounds included 2,4-dimethylpentane, 2-methylpropane, pentane, *cis*-2-pentene, *trans*-2-pentene, propane, hexane, butane, and 3-methylpentane. The similarity of their temporal patterns of production and elimination are also indicative that the same microbial metabolic pathways and mechanisms are being used to degrade the hydrocarbon components in the two microcosms.

The similarity between the degradation rates in the MFC and the SAM re-treated groups could also indicate that the two microcosm systems had evolved or been conditioned to become more functionally similar to each other (Figures 39-50). The sampling regime could have induced or directed the subsequent developments within the two systems so that their responses would be similar. The MFC and the SAM were both sampled twice a week using the same sampling device and the same techniques. The microcosms were stirred vigorously to re-suspend the detrital matter and sediments prior to the removal of the subsample. The re-suspension physically dispersed the clumps of blue-green algae, fragmented cells, fecal matter and sediment aggregates and served to re-expose more substrate surface area for utilization. Buried algal cells and other relatively high quality organic matter were also re-suspended to become available for further utilization. Each sampling event caused a pulsed release of nutritive organic material back into the system as well as exposing hydrocarbon absorbed and adsorbed substrates for photooxidative, volatilization, and microbial degradation processes. The pulsing of the organic matter, nutrients, and hydrocarbons twice a week probably helped to maintain certain organism concentrations, interactions, and rates of cycling for longer periods of time, than would have persisted without the impact of the sampling regimes.

The similarities in degradation rates between the MFC and the SAM microcosms may also be a result of the selective toxic effects of the jet fuels. Their effects could essentially re-structure the biological components so that above a certain water soluble concentration some microorganisms would always be eliminated or inhibited, while others would be activated or freed from competitive constraints. The resulting structural composition would be more uniform and consistent in the types of surviving microbial populations for each treatment group. These treatment selected microorganisms have the potential to determine and control the functional processes and responses for that treatment group. The faster rates of degradation for toluene, *m,p*-xylene, *o*-xylene, and ethylbenzene in the 5% WSF treatment groups in the MFC and the SAM may indicate that at those concentration levels the microbial utilization of these compounds was stimulated. At the 15% WSF treatment concentration the effects may be more inhibitory and caused a slower rate of utilization. The MFC 1% WSF treatment groups had the slowest rates of degradation that could indicate that the concentration levels of the aromatics were not at sufficient levels to induce the necessary enzymatic systems for utilization. It could also indicate that the concentrations of the aromatics were not sufficient to be preferred as a substrate for energy or as a carbon source (Alexander, 1985).

In support of the selective effects of toxic substances to microorganisms, previous studies using indigenous microbial populations exposed to various concentrations of organic toxicants in water, soil, sediment, and sewage have been conducted (Alexander, 1985). At low concentration levels the microbial populations mineralized the compounds to carbon dioxide. At intermediate concentration levels the toxicants were degraded by both mineralization and cometabolic mechanisms by the respective microbial communities. At high concentrations the microorganisms shifted to predominantly cometabolic pathways that produced organic intermediate compounds including alcohols, aldehydes, ketones, and carboxylic acids (Alexander, 1985). The explanations for this degradative pattern were that the mineralization of low levels of hydrocarbons may require oligotrophic microbial populations that are able to perform more enzymatically specialized oxidative processes. At the mid-range concentration levels the hydrocarbons were conducive for both types of metabolic pathways. At high concentration levels the absence of mineralization may be a result of the inhibition of mineralizing, but not the cometabolizing microbial populations (Alexander, 1985). These relationships would explain the similar degradation rate patterns in the same treatment groups when compared between the two microcosms. Similar shifts in microbial metabolic pathways and mechanisms for hydrocarbon degradation would occur

at the same percent WSF concentration levels, rather than at individual component concentration levels (Tables 15 and 16).

Hydrocarbon Degradation Dissimilarities

There were significant differences determined for the degradation rates of some hydrocarbon components both within each microcosm experiment and between microcosms. One explanation could involve the structural complexity of the hydrocarbon component being degraded. The oxidation of methyl- and alkyl-substituted aromatics require several enzymatic steps involving several types of specialized microorganisms. The alkyl side chain must first be oxidized by alkane degrading microorganisms followed by oxidation and cleavage of the aromatic ring structure that would involve several degradative steps. The more steps involved in the degradative process, the longer the process would take and the more opportunities there would be for differences in utilization rates to occur (Gibson, 1977). This would apply to the degradation of any hydrocarbon compound, but the significant differences that were determined did seem to be more prevalent for those compounds that rely on several oxidative steps, especially in the xylenes where the position of the methyl group on the aromatic ring will affect the ease of oxidation (Tables 11 and 18).

The initial structural and functional conditions in the MFC and the SAM microcosms were crucial in determining the hydrocarbon degradation rates in not only the initial treatment groups, but also in the re-treated groups. The differences in the degradation slopes of the hydrocarbon components in the initial microcosm experiments, compared to their respective slopes in the re-treated microcosms are obvious (Figures 19-30). In the re-treated MFC and SAM experiments the differences between the initial structural and functional conditions in the 0%+15% and the 15%+15% WSF treatment groups were much less obvious. The hydrocarbon degradation rates in the re-treated groups display their almost identical processes at the beginning of the re-treatment experiment. As the experiments progressed the component degradation slopes gradually begin to diverge. Some hydrocarbon components diverged to a greater extent than other components until they were significantly different from each other (Tables 17-19).

In the MFC the reference treatment microcosms (0% WSF) that were re-treated with the 15% WSF were able to degrade the hydrocarbon components faster than in the re-treated 15%+15% WSF treatment microcosms. In the SAM experiment it was the hydrocarbon components in the 15%+15% re-treated groups that were degraded at faster rates. If the initially treated 15% WSF groups had "recovered" to be comparable

to the untreated groups, then the two re-treated groups should not have diverged over time. The presumption would be that by the time of the second re-treatment event (sixty days after the initial treatment) both the 0% and the 15% WSF treatment groups would be "starting" from the same structural and functional state of development (Figures 39-50). If the relationship of faster degradation rates for pre-adapted microbial communities is valid, then the 15%+15% WSF treatment groups should have degradation rates that are at least as rapid as the 0%+15% WSF. The fact that they did diverge indicates that every treatment to a system has a subtle and lasting effect. These effects have the potential to affect and control the future structure and function of the system, that may not be detected or differentiated as an indirect result caused by a toxicant release using conventional analytical procedures.

Another factor that could account for the significant differences determined between the two microcosms was the length of time selected to compare the degradation slopes of the hydrocarbon components. The criteria used was the time required to mineralize the compound, or reduce the concentration to a minimum threshold level. If the compound was degraded completely the total length of time was used. If the compound decreased in concentration to a certain threshold level, the duration of time to reach the threshold level was used. If the compound was degraded to a certain level and then began to increase as a result of the production of the compound as a metabolite, the length of time selected was the interval of time to its increase. The selection of a shorter time interval for comparisons would have probably eliminated many of the significantly different determinations that would have emphasized those points on the regression that were the most similar (Figures 39-50). However, a shorter time interval would not have revealed subsequent degradative rate patterns.

Several other factors that may have contributed to the significant differences in hydrocarbon rate responses could have been the different size of the microcosms, the surface to volume ratios of the test chambers, and the quality and quantity of the detrital organic matter that would affect the partitioning and bioavailability of the hydrocarbon component for microbial utilization (Dewitt et al., 1992, Karickhoff et al., 1979). The MFC was smaller in size, but its surface to volume ratio was approximately one and one half times greater than the SAM microcosm. The adsorption of the hydrocarbons to the glass walls of the MFC microcosms would remove more organics from the water column and concentrate them for greater utilization by the microbial communities associated with the glass surfaces. The higher absorption and adsorption of the hydrocarbons to the higher molecular weight, high quality detrital organic matter in the MFC would also concentrate the compounds in microhabitats that have the greatest number of associated

microbial populations and would also be accessible for greater utilization. Conversely, the hydrocarbon components in the SAM would have less available glass surface area and organic matter on which to be adsorbed or absorbed. The hydrocarbons would partition between the organisms, the glass, the sediments, and the water column that might make the compounds less concentrated for utilization by the microorganisms.

The toxicity of the respective jet fuels could also have affected degradation rate responses. The high levels of aromatics in the JET-A WSF's confer greater immediate toxicological effects to the microcosm organisms compared to the less toxic alkanes in the JP-8 WSF. The decrease in the degradation rates in the JET-A MFC 15%+15% WSF treatment group could be attributed to the toxicity of the mixture. The initial 15% WSF treatment may have slightly inhibited microbial degradation, but the second 15% WSF treatment could have had an additive effect and inhibited degradation to a greater extent. Conversely, the JP-8 SAM WSF was less toxic and may have induced degradation by enabling more microbial populations to survive and adapt in the 15%+15% WSF treatments. An increase in the survival of the microorganisms could account for the slightly higher rates of degradation compared to the 0%+15% WSF treatment.

Microcosm Assessment

In the initial SAM experiment the rates of degradation "lagged" to the extent that they were significantly different from those in the later re-treatment experiment and from the degradative rates in the MFC experiments. The initial SAM experiment did display similarities in degradation rates for the individual components in each treatment group. It also displayed relatively consistent patterns in the ranking of the components by concentration and degradation rates in their alkane, aromatic or alkyl-aromatic chemical class. The SAM also displayed the shifts in the microbial metabolic pathways with changes in the water soluble fraction concentrations. The results were consistent between replicates and the ability to make decisions regarding the validity of a response or pattern observed could be made with some degree of confidence due to the standardization and the robustness of the statistical data.

The limitation of the initial SAM experiment was the initial conditions that caused most of the degradative response rates to be at significantly slower rates. The use of the SAM for risk estimations or for simple fate and effect studies may generate degradation rate studies that over-estimate the length of time the compound will be in the environment. Though this may be protective to the environment in discharge permitting, an accidental spill may encourage remediation techniques well beyond the immediate necessity warranted by the size of the release. A predictive ability that would

allow the appropriate placement of labor, finances and other resources to a spill site would be more protective to the environment and cost effective in the long term.

In the Mixed Flask Culture microcosm experiment the initial conditions generated degradative rates that were elevated when compared to all the other treatments and to the SAM experiments. This type of response has been quantified in previous laboratory experiments where the results obtained were due to the optimization of conditions that were not applicable to real ecosystem processes (Alexander, 1985; Gibson, 1974; Walker et al., 1976a). The MFC microcosms did display the same shifts in metabolic pathways as the concentrations of the water soluble fractions were increased. However, the results for the component degradation rates in the treatment groups were more variable and could have been dependent on a combination of factors affecting the initial microbial community in the microcosm. The inability to standardize and replicate these microcosms as closely as in the SAM protocol could account for much of the variability in the results.

The limitation of using the MFC initial microcosm protocol was the highly elevated degradative rate responses that in risk assessment studies could have underestimated the amount of time necessary to degrade the WSF mixture. The presumed shorter time interval for the degradation of the WSF mixture would cause greater damage to the environment by being under protective. Risk estimations would be based on degradative rates that were unrealistically elevated and could lead to the permissible discharge of hydrocarbon mixtures in excess of the ecosystem's capacity to degrade the compounds. Remediation attempts at a spill site may be understaffed or underfunded based on the assumption that the compounds would be rapidly removed by indigenous microbial populations within a short period of time.

Microcosm Validation

An assessment of the two microcosms as valid biological models of ecosystem structure and function were made in terms of their ability to display rate responses similar to those found in field studies. Increased rates in the degradation of the hydrocarbon components in those treatment groups that had been previously treated were expected, but only consistently observed in the SAM microcosm re-treatment experiment. The MFC microcosm re-treatment experiment results were more highly variable. The aromatics and alkyl-aromatics had consistently higher degradation rates in the 0%+15% WSF treatment group than in the 15%+15% WSF treatment (Table 16).

The Standardized Aquatic Microcosm does seem to be the most consistent,

replicable, reproducible, and valid microcosm model for use in ecosystem-level studies involving complex mixtures of hydrocarbons. The microcosm dynamics allowed the determination of subtle changes in microbial degradative metabolic pathways and were comparable to field determined results. The weakness of using the SAM protocol is the dynamic role that the Daphnia play in the structural and functional responses in the microcosms. Their presence or absence can determine the evolutionary direction of the microcosm system after the addition of the toxicant. The microcosm experiment becomes a test of the sensitivity of the Daphnia to the toxicant that can determine the sensitivity of the entire system to the toxicant (Sugiura, 1992). The other weakness is the construction and use of a relatively sterile initial system. The low species diversity constrains the system to a degree by placing too great of importance on the responses of very few representative species that are often not well researched. In addition, the lack of established and interactive microbial, detritivore, and macroinvertebrate communities in conjunction with active energy and nutrient cycling, severely limits the types of possible structural and functional responses to the toxicant effects. Rather than a "sensitized" system the initial SAM microcosms are handicapped in the types of rate responses and biological sublethal endpoints that can be measured.

Some recommendations would be to increase the period of acclimation prior to treatment from seven days to possibly fourteen, to replace the *Daphnia magna* with a less efficient grazer such as a *Ceriodaphnia* species, and increase the types of detritivore and microorganisms present. Sampling regimes and analyses must be modified to include greater sampling frequencies when conducting any studies involving microbial communities. Volatile hydrocarbon components that are produced from the degradation of biogenically derived aromatic compounds could also be included as a parameter to measure. The elevated production of these compounds after treatment may be used as an indicator of toxicant stress to the microbial community. Whether these releases would be correlated to other types of stress events occurring in natural environments would have to be investigated.

Some parameters that should not be included in the monitoring and measurement of microbial degradative mechanisms and metabolic rates are carbon dioxide and radiolabeled biomarkers. Alexander (1985) found that the exposure of indigenous microbial populations to low concentrations of hydrocarbons stimulated the release of carbon dioxide. The release of the carbon dioxide was not a stress related response by the microorganism, but was due to the metabolic pathways used by the microorganisms to mineralize the hydrocarbons to inorganic carbon. McKeena and Kallio (1964) reported that alkane hydrocarbons increased the oxygen consumption and respiratory

release of carbon dioxide in microbial populations, but there was no concomitant oxidative degradation of the hydrocarbons. The elevated release of carbon dioxide that is presumed to be a significant functional response was found to be in this situation merely respiratory stimulation. Alternatively, the use of radiolabeled genetic or enzymatic markers in long term microbial studies are severely limited by the rapid generation times of the microorganisms. Within a short period of time of two to three days the radiolabel can be "diluted" to below detection limits in the resulting progeny (Atlas and Bartha, 1994).

Conclusions

The MFC and the SAM microcosms have both strengths and weaknesses in their experimental designs. It becomes the responsibility of the investigator to evaluate these systems and determine the appropriateness of the microcosm type for the proposed research. Both microcosms were developed as a basic protocol with the flexibility to be manipulated or modified to meet the needs of the researcher (Taub, 1984). The importance of having this flexibility is that it allows the investigator to focus the research on the time dependent changes in population densities and metabolic processes relevant to the hypothesis being tested. It is these processes that are the important indicators of the effects of a toxicant on ecosystem function and not the actual structural composition within the microcosms (Sugiura, 1992). The research hypothesis should never be formulated to conform within the constraints defined by the microcosm protocol.

The hydrocarbon degradation rates in the initial MFC and the SAM WSF treatment groups were significantly different from each other for most hydrocarbon components. The initial conditions that existed in each of the microcosms were responsible for these differences, rather than the concentrations or compositions of the water soluble fractions they were treated with. The concentrations of the specific hydrocarbon class of compounds in the WSF will determine the types of microbial populations degrading the components, not the concentrations of the individual hydrocarbon components. The rates at which the individual hydrocarbons are degraded will be dependent on the types of functionally active microorganisms, the initial structural and functional conditions in the systems, and the chemical structure and properties of the hydrocarbon.

The microorganisms will degrade each hydrocarbon class of compounds using specific enzymatic mechanisms and metabolic pathways that are consistent, regardless of the actual composition of the microbial community or the composition of the jet fuel. The metabolic by-products and intermediates formed during the hydrocarbon degradative processes were almost identical in the two microcosms and reflect the similarity in functional processes of the two microbial communities. The MFC and the SAM microbial communities also displayed similar shifts in metabolic pathways and mechanisms for hydrocarbon degradation that were dependent on the concentration of the water soluble fractions rather than individual component concentration levels.

The major hydrocarbon components in the water soluble fractions in both microcosms were mineralized within a period of approximately two weeks. Subsequent degradations of higher molecular weight, polycyclic aromatic compounds occurred later

in both the MFC and the SAM. Lower molecular weight aromatic and alkane metabolites were released into the water column and detected for up to one month after treatment. The presence of these metabolites as well as metabolites from the degradation of biogenic aromatic compounds were perceived as indirect treatments to the systems. The concentration of these treatments were low in comparison to the initial treatment of the WSF of the jet fuel, but could indirectly serve to influence the structural and functional processes in the microcosms.

The ecosystem implications from these microcosm tests are that the disappearance of a toxicant from one or a few of the parameters measured in the system does not imply that the perturbation to the system is finished and that the system has recovered. The indirect effects caused by chronic low level releases of biogenically produced toxic metabolites, bioaccumulated toxicants from dying organisms and lysed cellular material, anthropogenic activities, and natural perturbations can potentially control and modify structural and functional processes. These indirect effects can be as subtle as a small change in the cycling rate of a specific nutrient or in the rate of nitrogen fixation in photoautotrophic bacteria that will have profound long term effects on the functional integrity of the ecosystem. The diversity of microhabitats that exist in every ecosystem can potentially retain and periodically release these toxicant or their intermediates for years or decades after the initial toxicant release. The effects generated by these releases will extend far beyond that interval of time.

The results from this study indicate that the initial conditions within the microcosms will determine the dynamics within the system long after the perturbation or toxicant is gone. The degradation rates in the MFC and the SAM became more similar after the microcosms had acclimated and matured during the two months prior to the re-treatment. The divergence of the degradation rates between the two re-treated groups in both the MFC and the SAM indicate that though the two systems were treated with the same concentration of the water soluble fraction of a jet fuel the prior history of the systems eventually influenced their future functional rate responses. These altered rate responses were initially indistinguishable from the reference re-treated groups, but during the course of the experiment they eventually caused the two groups to display distinctly different degradation rates. In the MFC the divergence was caused by a decrease in degradative capacity of the microbial community, while in the SAM the divergence was caused by an increase in microbial degradative ability. The replication of the divergence in the rate responses in the two microcosms suggests that ecosystems will respond similarly after the initial toxicant exposure event. However, the eventual rate that the system will degrade the toxicant and the subsequent effects on the microbial

and higher organism structure and functional processes will be dependent on, and a culmination of, that system's ecological history.

The validity of the SAM and the MFC microcosm as biological models of ecosystem structure and function was tested on the criterion of their ability to display increased rates of degradation in the re-treated groups similar to observed responses in field studies. The SAM did display increased rates of degradation for almost all hydrocarbon components in the re-treatment experiment and are considered valid generic simulations of ecosystem processes and dynamics (Table 16). The SAM's have the added advantages of being replicable and reproducible, producing consistent and statistically robust results. The SAM's are the preferred microcosm model for this type of study. Conversely, the MFC displayed almost no increased degradation rates in the re-treated microcosm experiment. In addition, though they could produce results comparable to the SAM's, any interpretation of the results would have been difficult due to the high variability in the data.

Whether all ecosystems display the same patterns and behaviors in their structural and functional relationships and processes that can be utilized in research to extrapolate results to the ecosystem-level of organization is still not clearly defined. The similarities determined between the two microcosms in their rates of degradation that were independent of the microcosm type, jet fuel composition, species diversity and trophic level complexity could imply that there existed universal ecosystem properties and universal patterns of responses to stress. However, the similarities could also merely be a function of the chemical structure of the hydrocarbon component that is degraded in rate controlled reactions that are catalyzed by an initial, microbially-mediated oxidizing event.

References

Aelion, C.M., D.C. Dobbins, and F.K. Pfaender. 1989. Adaptation of aquifer microbial communities to the biodegradation of xenobiotic compounds: Influence of substrate concentration and preexposure. *Environ. Toxocol. Chem.* 8:75-86.

Alexander, M. 1985. Biodegradation of organic chemicals. *Environ. Sci. Technol.* 18:106-111.

ASTM D2887. 1986. Boiling range distribution of petroleum fractions by gas chromatography. In *1986 Annual Book of ASTM Standards.* Vol. 5.02, pp. 658-667.

ASTM D3710. 1986. Boiling range distribution of gasoline and gasoline fractions by gas chromatography. In *1986 Annual Book of ASTM Standards.* Vol. 5.03, pp. 99-113.

ASTM E1366-91. 1991. Standard practice for standardized aquatic microcosms: fresh water. In *1991 Annual Book of ASTM Standards.* Vol. 11.04, pp. 1017-1051.

Atlas, R.M. 1975. Effects of temperature and crude oil composition on petroleum biodegradation. *Appl. Microbiol.* 30:396-403.

Atlas, R.M. 1978. Microorganisms and petroleum pollutants. *Bioscience* 28:387-391.

Atlas, R.M. 1981. Microbial degradation of petroleum hydrocarbons: an environmental perspective. *Microbial Rev.* 45:180-209.

Atlas, R.M. 1988. Biodegradation of hydrocarbons in the environment. In G.S. Omenn, ed., *Environmental Biotechnology, Reducing Risks from Environmental Chemicals Through Biotechnology.* Plennum Press, New York, NY, pp. 211-222.

Atlas, R.M. 1991. Microbial hydrocarbon degradation - Bioremediation of oil spills. *J. Chem. Tech. Biotechnol.* 52:149-156.

Atlas, R.M. and G.S. Sayler. 1988. Tracking microorganisms and genes in the environment. *Environmental Biotechnology, Reducing Risks from Environmental Chemicals Through Biotechnology.* G.S. Omenn, Editor. Plennum Press, New York. pp. 31-49

Atlas, R.M. and R. Bartha. 1973. Inhibition by fatty acids of the biodegradation of petroleum. *Antonie van Leeuwenhoek* 39:257-271.

Atlas, R.M. and R. Bartha. 1992. Hydrocarbon biodegradation and oil spill bioremediation. In K.C. Marshall ed., *Advances in Microbial Ecology, Volume 12.* Plenum Press, New York, NY, pp. 287-338.

Atlas, R.M. and R. Bartha. 1993. *Microbial Ecology: Fundamentals and Applications, Third Edition.* Benjamin/Cummings Publishing Co., Menlo Park. 563pp.

Atlas, R.M., A. Horowitz and M. Busdosh. 1978. Prudhoe crude oil in arctic marine ice, water, and sediment ecosystems: Degradation and interactions with microbial and benthic communities. *J. Fish. Board Can.* 35:585-590.

Atlas, R.M., A. Horowitz, M. Krichevsky and A.K. Bej. 1991. Response of microbial populations to environmental disturbance. *Microb. Ecol.* **22**:249-256.

Bailey, N.J.L., A.M. Jobson and M.A. Rogers. 1973. Bacterial degradation of crude oil: Comparison of field and experimental data. *Chemical Geology.* 11:203-221.

Baker, J.M. 1971. Studies on salt marsh communities. In E.B. Cowell, Ed., *The Ecological Effects of Oil Pollution on Littoral Communities.* Elsevier Publishing Co. Ltd., New York, NY, pp. 16-101.

Baker, J.M. 1971. The Ecological Effects Of Oil Pollution on Littoral Communities. E.B. Cowell, Editor Institute of Petroleum. Elsevier Publishing Co. Ltd., London, England. pp. 88-98.

Bartell, S.T., R.H. Gardner and R.V. O'Neill. 1992. *Ecological Risk Estimation.* Lewis Publishers, Chelsea, MI. pp252.

Beer, J.V. 1968. Post-mortem findings in oiled auks dying during attempted rehabilitation. In J.D. Carthy and D.R. Arthur, eds., *The biological Effects of Oil Pollution on Littoral Communities.* Field Studies Council, London, Eng., pp. 123-129.

Berry, W.O. and J.D. Brammer. 1977. Toxicity of water-soluble gasoline fractions to fourth-instar larvae of the mosquito *Aedes aegypti* L. Environ. Pollut. 13:229-234.2

Beyers, R.J. 1963. The metabolism of twelve laboratory microecosystems. *Ecol. Monogr.* **33**:281-306.

Bird, C.W. and P. Molton. 1969. The biochemical status of metabolites of alkane-utilizing *pseudomonas* organisms. *Biochem. J.* **114**:881-884.

Boesch, D.F., C.H. Hershner and J.H. Milgram. 1974. *Oil Spills and the Marine Environment.* Ballinger Publishing Company, Cambridge, MA. 114pp.

Bourne, W.R.P. 1968. Oil pollution and bird populations. In J.D. Carthy and D.R. Arthur, eds., *The biological Effects of Oil Pollution on Littoral Communities.* Field Studies Council, London, Eng., pp. 99-121.

Branson, D.R. 1978. Predicting the fate of chemicals in the aquatic environment from laboratory data. In J. Cairns, Jr., K.L. Dickson and A.W. Maki, eds., *Estimating the Hazard of Chemical Substances to Aquatic Life, ASTM STP 657.* American Society for Testing and Materials, Philadelphia, PA, pp. 55-70.

Brock, T.D., M.T. Madigan, J.M. Martinko and J. Parker. 1994. *Biology of Microorganisms, Seventh Edition.* Prentice-Hall, Inc., Englewood Cliffs, NJ, pp. 623-691.

Brown, W.H. 1982. *Third Edition Introduction to Organic Chemistry.* Willard Grant Press, Boston, MA, pp. 328a-328e.

Browning, E. 1953. *Toxic Solvents.* Edward Arnold & Co., London, Eng., pp. 1-41.

Bull, A.T. and J.H. Slater. 1982. *Microbial Interactions and Communities. Volume 1.* Academic Press Inc., London, England. pp. 13-44.

Butler, J.N. and E.M. Levy. 1978. Session I. summary and overview. log-term fate of petroleum hydrocarbons after spills-compositional changes and microbial degradation. *J. Fish. Res. Board Can.* **35**:604-605.

Cerniglia, C.E., D.T. Gibson and C. Van Baalen. 1979. Algal oxidation of aromatic hydrocarbons: formation 1-naphthol from naphthalene by *agmenellum quadruplicaum*, strain PR-6. *Biochem. Biophys. Res. Comm.* **88**:50-58.

Connell, D.E. and G.J. Miller. 1984. *Chemistry and Ecotoxicology of Pollution.* John Wiley & Sons, Inc., New York, NY, pp. 228-249.

Cote, R.P. 1976. *The Effects of Petroleum Refinery Liquid Wastes on Aquatic Life, With Special Emphasis on the Canadian Environment.* National Research Council Canada. Associate Committe on Scientific Criteria for Environmental Quality. NRCC No. 15021. Publications, NRC, Ottawa. 77pp.

Davies, J.A. and D.E. Hughes. 1968. The biochemistry and microbiology of crude oil degradation. In J.D. Carthy and D.R. Arthur, eds., *The Biological Effects of Oil Pollution on Littoral Communities.* Field Studies Council, London, Eng., pp. 139-144.

Davis, J.B. 1967. *Petroleum Microbiology.* Elsevier Publishing Company, New York, NY. 604pp.

Delille, D. and R. Siron. 1993. Effect of dispersed oil on heterotrophic bacterial communities in cold marine waters. *Microb. Ecol.* **25**:263-273.

Dewitt, T.H., R.J. Ozretich, R.C. Swartz, J.O. Lamberson, D.W. Schults, G.R. Ditsworth, J.K.P. Jones, L. Hoselton and L.M. Smith. 1992. The influence of organic matter on the toxicity and partitioning of sediment-associated fluoranthene. *Environ. Toxicol. Chem.* **11**:197-208.

Dibble, J.T. and R. Bartha. 1979. Effects of environmental parameters on the biodegradation of oil sludge. *Appl. Environ. Microbiol.* **37**:729-739.

Donahue, W.H., R.T. Wang, M. Welch and J.A.C. Nicol. 1977. Effects of water-soluble components of petroleum oils and aromatic hydrocarbons on barnacle larvae. *Environ. Pollut.* **13**:187-202.

Evans, P.J., D.T. Mang and L.Y. Young. 1991. Degradation of toluene and m-xylene and transformation of o-xylene by denitrifying enrichment cultures. *Appl. Environ. Microbiol.* **57**:450-454.

Fedorak, P.M. and D.W.S. Westlake. 1981. Microbial degradation of aromatics and saturates in prudhoe bay crude oil as determined by glass capillary gas chromatography. *Can. J. Microbiol.* **27**:432-443.

Fedorak, P.M. and D.W.S. Westlake. 1984. Microbial degradation of alkyl carbazoles in norman wells crude oil. *Appl Environ. Microbiol.* **47**:858-862.

Fenchel, T. 1977. The significance of bactivorous protozoa in the microbial community of detrital particles. In J. Cairns, Jr. ed., *Aquatic Microbial Communities.* Garland Publishing, Inc., New York, NY, pp. 529-544.

Focht, D.D. 1988. Performance of biodegradative microorganisms is soil: Xenobiotic Chemicals as unexploited metabolic niches. In G.S. Omenn, ed., *Environmental Biotechnology, Reducing Risks from Environmental Chemicals Through Biotechnology.* Plennum Press, New York, NY, pp. 15-29.

Focht, D.D. and D.W. Westlake. 1988. Degradation of polycyclic aromatic hydrocarbons and aromatic heterocycles by a *pseudomonas* species. *Can J. Microbiol.* **34**:1135-1141.

Focht, D.D., P.M. Fedorak and D.W. Westlake. 1990. Mineralization of [^{14}C]hexadecane and [^{14}C]phenanthrene in crude oil: Specificity among bacterial isolates. *Can J. Microbiol.* **36**:169-175.

Ford, T.E. 1993. *Aquatic Microbiology An Ecological Approach.* Blackwell Scientific Publications, Oxford, Eng., pp. 1-14.

Fredericks, K.M. 1966. Adaptation of bacteria from one type of hydrocarbon to another. *Nature* **209**:1047-1048.

Gibson, D.T. 1977. Biodegradation of aromatic hydrocarbons. In D.A. Wolfe, ed., *Fate and Effects of Petroleum Hydrocarbons in Marine Ecosystems and Organisms.* Proceedings of a Symposium, November 10-12, 1976, Olympic Hotel, Seattle, Washington. Pergamon Press, New York, NY, pp. 36-46.

Gibson, D.T. 1978. Microbial transformation of aromatic pollutants. In O. Hutzinger and S. Safe, Eds., *Volume 1. Aquatic Pollutants: Transformations and Biological Effects.* Pergamon Press Ltd., New York, NY pp. 187-204.

Gibson, D.T., G.E. Cardini and R.E. Kallio. 1970. Incorporation of oxygen-18 into benzene by *pseudomonas putida. Biochemistry.* **9**:1631

Gibson, D.T., V. Mahadevan and J.F. Davey. 1974 Bacterial metabolism of *para*- and *meta*-xylene: Oxidation of the aromatic ring. *J. Bacteriol.* **119**:930-936.

Giddings, J.M. 1983. Microcosms for assessment of chemical effects on the properties of aquatic ecosystems. In J. Saxena, ed., *Hazard Assessment of Chemicals. Current Developments, Vol. 2.* Academic Press, Inc., New York, NY, pp. 45-94.

Hammons, A.S., J.M. Giddings, G.W. Suter II, L.W. Barnthouse and J.V. Nabholz. 1981. *Methods for Ecological Toxicity. A Critical Review of Laboratory Multispecies Tests.* Ann Arbor Science Publishers, Inc., Ann Arbor, MI, pp. 1-164.

Harrass, M.C. and P.G. Sayre. 1989. Use of microcosm data for regulatory decisions. In U.M. Cowgill and L.R. Williams, eds., *Aquatic Toxicology and Hazard Assessment: Twelfth Volume, ASTM STP 1027.* American Society for Testing and Materials, Philadelphia, PA, pp. 204-223.

Horowitz, A. and R.M. Atlas. 1977a. Continuous open flow-through system as a model for oil degradation in the arctic ocean. *Appl. Environ. Microbiol.* **33**:647-653.

Horowitz, A. and R.M. Atlas. 1977b. Response of microorganisms to an accidental gasoline spillage in an arctic freshwater ecosystem. *Appl. Environ. Microbiol.* **33**:1252-1258.

Horvath, R.S. 1972. Microbial co-metabolism and the degradation of organic compounds in nature. *Bacteriological Reviews* **36**:146-155.

Horvath, R.S. and M. Alexander. 1970. Cometabolism: a technique for the accumulation of biochemical products. *Can. J. Microbiol.* **15**:1131-1132.

Hou, C.T. 1982. Microbial transformation of important industrial hydrocarbons. In J.P. Rosazza, ed., *CRC Microbial Transformations of Bioactive Compounds. Volume I.* CRC Press, Inc., Boca Raton, FL, pp. 81-107.

Howarth, R.W. 1989. Determining the ecological effects of oil pollution in marine ecosystems. In S.A. Levin, M.A. Harwell, J.R. Kelly, and K.D. Kimball, eds., *Ecotoxicology: Problems and Approaches.* Springer-Verlag, New York, NY, pp. 69-97.

Hutchins, S.R. 1991. Biodegradation of monoaromatic hydrocarbons by aquifer microorganisms using oxygen, nitrate or nitrous oxide as the terminal electron acceptor. *Appl. Environ. Microbiol.* **57**:2403-2407.

Jamison, V.W., R.L. Raymond and J.O. Hudson. 1976. Biodegradation of high-octane gasoline. In J.M. Sharpley and A.M. Kaplan, eds., *Proceedings of the Third International Biodegradation Symposium.* Applied Science Publishers, London, Eng., pp. 187-196.

Jobson, A., F.D. Cook and D.W.S. Westlake. 1972. Microbial utilization of crude oil. *Appl. Microbiol.* **23.**:1082-1089.

Johnson, M.J. 1964. Utilization of hydrocarbons by micro-organisms. *Chemistry and Industry.* pp. 1532-1537.

Kampfer, P., M. Steiof, and W. Dott. 1991. Microbial characterization of a fuel-oil contaminated site including numerical identification of heterotrophic water and soil bacteria. *Microbial Ecol.* **21**:227-251.

Kappeler, T. and K. Wuhrmann. 1978. Microbial degradation of the water-soluble fraction of gas oil-I. *Water Research* **12**:327-333.

Kappeler, T. and K. Wuhrmann. 1978. Microbial degradation of the water-soluble fraction of gas oil-II bioassays with pure strains. *Water Research* **12**:335-342.

Karickhoff, S.W. 1984. Organic pollutant sorption in aquatic systems. *J. Hydraulic Engineering.* **110**:707-735.

Karickhoff, S.W., D.S. Brown and T.A. Scott. 1979. Sorption of hydrophobic pollutants on natural sediments. *Water Research.* **13**:241-248.

Kauss, P.B. and T.C. Hutchinson. 1975. The effects of water-soluble petroleum components on the growth of *Chlorella vulgaris* Beijerinck. *Environ. Pollut.* **9**:157-174.

Keith, L.H. 1974. Chemical characterization of industrial waste waters by gas chromatography-mass-spectrometry. *Sci. Total Environ.* **3**:87-102.

Kester, A.S. and J.W. Foster. 1963. Diterminal oxidation of long-chain alkanes by bacteria. *J. Bacteriology.* **85**:859-869.

Kindig, A.C., L.L. Conquest and F.B. Taub. 1983. Differential sensitivity of new versus mature synthetic microcosms to streptomycin sulfate treatment. In W.E. Bishop, R.D. Cardwell and B.B. Heidolph, eds., *Aquatic Toxicology and Hazard Assessment: Sixth Symposium, ASTM 802.* American Society for Testing and Materials, Philadelphia, PA, pp. 192-203.

Knap, A.H. and P.J. Williams. 1981. Experimental studies to determine the fate of petroleum hydrocarbons from refinery effluent on an estuarine system. *Environ. Sci. Technol.* **16**:1-3.

Kroer, N. and R.B. Coffin. 1992. Microbial trophic interactions in aquatic microcosms designed for testing genetically engineered microorganisms: A field comparison. *Microbial Ecol.* **23**:143-157.

Kroer, N. R.B. Coffin and N.O.G. Jorgensen. 1994. Comparisons of microbial trophic interactions in aquatic microcosms designed for the testing of introduced microorganisms. *Environ. Toxicol. Chem.* **13**:247-257.

Kuhn, E.P., P.J. Colberg, J.L. Schnoor, O. Wanner, A.J.B. Zehnder, and R.P. Schwarzzenbach. 1985. Microbial transformations of substituted benzenes during infiltration of river water to groundwater: Laboratory column studies. *Environ. Sci. Technol.* **19**:961-968.

Landis, W.G., R.A. Matthews, A.J. Markiewicz, N.A. Shough and G.B. Matthews. 1993. Multivariate analyses of the impacts of the turbine fuel JET-A using a microcosm toxicity test. *J. Environ. Sci.* **2**:113-130.

Landis, W.G., A.J. Markiewicz, R.A. Matthews, M.J. Roze and G.B. Matthews. 1994. Comparison of the community dynamics of the standardized aquatic microcosm and mixed flask culture microcosm toxicity tests with suggestions for design criteria in multispecies toxicity testing. In Press.

Leffler, J.W. 1980. In J.P. Giesy, ed., *Microcosms in Ecological Research.* U.S. Dept of Energy, Symposium Series 52, CONF-761101. Washington, DC, pp. 14-29.

Manahan, S.E. 1991. *Environmental Chemistry Fifth Edition.* Lewis Publishers, Chelsea, Michigan. 583pp.

Manahan, S.E. 1989. *Toxicological Chemistry, A Guide to Toxic Substances in Chemistry.* Lewis Publishers Inc., Chelsea, MI. 317pp.

Matthews, R.A., A.L. Buikema Jr., J. Cairns Jr. and J.H. Rodgers Jr. 1982. Biological monitoring. Part IIB-receiving system functional methods, relationships and indices. *Water Res.* **16**:129-139.

McAuliffe, C. 1966. Solubility in water of paraffin, cycloparaffin, olefin, acetylene, cycloolefin, and aromatic hydrocarbons. *J. Physical Chemistry.* **70**:1267-1275.

McKenna, E.J. 1977. Biodegradation of polynuclear aromatic hydrocarbon pollutants by soil and water microorganisms. UILU-WRC-76-0113, University of Illinois, Urbana Water Resources Center, Urbana, IL. 25pp.

Moore, J.W. and S. Ramamoorthy. 1984. *Organic Chemicals in Natural Waters. Applied Monitoring and Impact Assessment.* Springer-Verlag, New York, NY, pp. 43-66, and 115-140.

Nalewajko, C. 1977. Extracellular release in freshwater algae and bacteria: Extracellular products of algae as a source of carbon for heterotrophs. In J. Cairns, Jr. ed., *Aquatic Microbial Communities.* Garland Publishing, Inc., New York, NY, pp. 589-624.

O'Neill, R.V. and J.B. Waide. 1981. Ecosystem theory and the unexpected: Implications for environmental toxicology. In B.W. Cornby, ed., *Management of Toxic Substances in Our Ecosystems. Taming the Medusa.* Ann Arbor Science Publishers, Inc., Ann Arbor, MI, pp. 43-73.

O'Neill, R.V. and J.M. Giddings. 1979. Population interactions and ecosystem function: Phytoplankton competition and community production. In G.S. Innis and R.V. O'Neill, eds., *Systems Analysis of Ecosystems.* International Co-operative Publishing, Fairland, MD, pp. 103-123.

Ooyama, J. and J.W. Foster. 1965. Bacterial oxidation of cycloparaffinic hydrocarbons. *Antonie van Leeuwenhoek.* **31**:45-65.

Perry, J.A. and N.H. Troelstrup. 1988. Whole ecosystem manipulation: A productive avenue for test system research? *Environ. Toxicol. Chem.* **7**:941-951.

Perry, J.J. 1979. Microbial cooxidations involving hydrocarbons. *Microbiol. Rev.* **43**:59-72.

Perry, J.J. 1980. Oil in the biosphere. In F.E. Guthrie and J.J. Perry, ed., *Introduction to Environmental Toxicology.* Elsevier Press, New York, NY, pp. 198-209.

Pirnik, M.P., R.M. Atlas and R. Bartha. 1974 Hydrocarbon metabolism by *brevibacterium erythrogenes:* normal and branched alkanes. *J. Bacteriology.* **119**:868-878.

Portier, R.J. 1985. Comparison of environmental effect and biotransformation of toxicants on laboratory microcosm and field microbial communities. *Validation and Predictability of Laboratory Methods for Assessing the Fate and Effects of Contaminants in Aquatic Ecosystems, ASTM STP 865.*, T.P. Boyle, Editor. American Society for Testing and Materials, Philadelphia, PA, pp. 14-30.

Pratt, J.R., N.J. Bowers, B.R. Niederlehner and J. Cairns, Jr. 1988. Symposium. Population and community level responses to environmental perturbations. Effects of chlorine on microbial communities in naturally derived microcosms. *Environ. Toxicol. Chem.* **7**:679-687.

Rappoport, Z. 1967. *CRC Handbook of Tables for Organic Compound Identification, Third Edition.* CRC Press, Inc., Boca Raton, FL, pp. 35.

Riser-Roberts, E. 1992. *Bioremediation of Petroleum Contaminated Sites.* CRC Press, Inc., Boca Raton, FL, pp. 1-197, A-i-A264.

Robinson, D.S. 1964. Oxidation of selected alkanes and related compounds by a *pseudomonas* strain. *Antonie van Leeuwenhoek.* 30:303-316.

Rochkind, M.L., J.W. Blackburn, and G.S. Sayler. 1986. *Microbial Decomposition of Chlorinated Hydrocarbons.* U.S. Environmental Protection Agency, Cincinnati, OH. EPA-600/2-86-090. 269pp.

Saunders, G.W. 1977. Carbon flow in the aquatic system. In J. Cairns, Jr. ed., *Aquatic Microbial Communities.* Garland Publishing, Inc., New York, NY, pp. 417-440.

Schaeffer, T.L., S.G. Cantwell, J.L. Brown, D.S. Watt and R.R. Fall. 1979. Microbial growth on hydrocarbons: terminal branching inhibits biodegradation. *Appl. Environ. Microbiol.* 38:742-746.

Shannon, L.J. and R.L. Anderson. 1989. Use of the mixed flask culture (mfc) microcosm protocol to estimate the survival and effects of microorganisms added to freshwater ecosystems. U.S. Environmental Protection Agency, Environmental Research Lab.-Duluth, MN. EPA PB-89-221295.

Shannon, L.J., M.C. Harrass, J.D. Yount and C.T. Walbridge. 1986. A comparison of mixed flask culture and standardized laboratory model ecosystems for toxicity testing. In J. Cairns, Jr., ed., *Community Toxicity Testing, ASTM STP 920.* American Society for Testing and Materials, Philadelphia, PA, pp. 135-157.

Shannon, L.J., T.E. Flum, R.L. Anderson and J.D. Yount. 1989. Adaptation of mixed flask culture microcosms for testing the survival and effects of introduced microorganisms. In U.M. Cowgill and L.R. Williams, eds., *Aquatic Toxicology and Hazard Assessment: Twelfth Volume, ASTM STP 1027.* American Society for Testing and Materials, Philadelphia, PA, pp. 224-239.

Sheehan, P.J. 1984a. Effects on community and ecosystem structure and dynamics. In P.J. Sheehan, D.R. Miller, G.C. Butler and P. Bourdeau, eds., *Effects Of Pollutants at the Ecosystem Level.* John Wiley and Sons Publishers, New York, NY, pp. 51-99.

Sheehan, P.J. 1984b. Functional changes in the ecosystem. In P.J. Sheehan, D.R. Miller, G.C. Butler and P. Bourdeau, eds., *Effects Of Pollutants at the Ecosystem Level.* John Wiley and Sons Publishers, New York, NY, pp. 51-99.

Sheehan, P.J. 1989. Statistical and nonstatistical considerations in quantifying pollutant-induced changes in microcosms. In U.M. Cowgill and L.R. Williams, eds., *Aquatic Toxicology and Hazard Assessment: Twelfth Volume, ASTM STP 1027.* American Society for Testing and Materials, Philadelphia, PA, pp. 178-188.

Stay, F.S., A. Katko, C.M. Rohm, M.A. Fix and D.P. Larson. 1988. Effects of fluorene on microcosms developed from four natural communities. *Environ. Toxicol. Chem.* 7:635-644.

Stay, F.S., T.E. Flum, L.J. Shannon and J.D. Yount. 1989. An assessment of the precision and accuracy of SAM and MFC microcosms exposed to toxicants. In U.M Cowgill and L.R. Williams, eds., *Aquatic Toxicology and Hazard Assessment: Twelfth Volume, ASTM STP 1027.* American Society for Testing and Materials, Philadelphia, PA, pp. 189-203.

Sugiura, K. 1992. A multispecies laboratory microcosm for screening ecotoxicological impacts of chemicals. *Environ. Toxicol. Chem.* **11**:1217-1226.

Suter II, G.W. 1993. *Ecological Risk Assessment.* Lewis Publishers, Chelsea, MI. 538pp.

Swindoll, C.M., S. Landon-Arnold and D.I. Flynn. 1989. Characterization of indigenous microbial degraders of aviation fuel components from a contaminated site. In Y.C. Yu, ed., *International Conference on Physiochemical and Biological Detoxification of Hazardous Wastes, Volume Two / Session VIII-Session XII.* Technomic Publishing Co. Inc., Lancaster, PA, pp. 977-991.

Taub, F.B. 1984. Synthetic microcosms as biological models of algal communities. In L.E. Shubert, ed., *Algae As Ecological Indicators.* Academic Press, Inc., London, Eng., pp. 363-394.

Taub, F.B. and M.E. Crow. 1980. Synthesizing aquatic microcosms. In J.P. Giesy, ed., *Microcosms in Ecological Research.* U.S. Dept. of Energy Symposium Series 52, CONF-761101. Washington, DC, pp. 69-104.

Taub, F.B. and P.L. Read. 1982. Model ecosystems: design, development, construction and testing, Vol. II: Standard aquatic microcosm protocol. In FDA Contract Report No. 223-80-2352. Food and Drug Administration, Washington, DC, pp. 1-55.

Taub, F.B., P.L. Read, A.C. Kindig, M.C. Harrass, H.J. Hartman, L.L. Conquest, F.J. Hardy and P.T. Munro. 1983. In W.E. Bishop, R.D. Cardwell and B.B. Heidolph, eds., *Aquatic Toxicology and Hazard Assessment: Sixth Symposium, ASTM 802. American Society for Testing and Materials, Philadelphia, PA, pp. 5-25.*

Thomas, J.M., J.R. Yordy, J.A. Amador and M. Alexander. Rates of dissolution and biodegradation of water-insoluble organic compounds. *Appl. Environ. Microbiol.* **52**:290-296.

Timmis, K.N., F. Rojo and J.L. Ramos. 1988. Prospects for laboratory engineering of bacteria. In G.S. Omenn, ed., *Environmental Biotechnology, Reducing Risks from Environmental Chemicals Through Biotechnology.* Plennum Press, New York, NY, pp. 61-79.

USEPA. 1979. *Methods for Chemical Analysis of Water and Wastes.* U.S. Environmental Protection Agency, Cincinnati, OH. EPA-600/4-79-020.

USEPA. 1982a. Purgeable halocarbons-method 601. In J.E. Longbottom and J.J. Lichtenberg, eds., *Methods for Organic Chemical Analysis of Municipal and Industrial Wastewater.* U.S. Environmental Protection Agency, Cincinnati, OH. EPA-600/4-82-057, pp. 601-1 - 601-10.

USEPA. 1982b. Purgeable aromatics-method 602. In J.E. Longbottom and J.J. Lichtenberg, eds., *Methods for Organic Chemical Analysis of Municipal and Industrial Wastewater.* U.S. Environmental Protection Agency, Cincinnati, OH. EPA-600/4-82-057, pp. 602-1 - 602-9.

Van Eyk, J. and T.J. Bartels. 1968. Paraffin oxidation in *Pseudomonas aeruginosa.* I. induction of paraffin oxidation. *J. Bacteriol.* **96**:706-712.

Vandermeulen, J.H. 1978. Introduction to the symposium on recovery potential of oiled marine northern environments. *J. Fish. Res. Board Can.* **35**:505-508.

Vandermeulen, J.H. and T.P. Ahern. 1976. Effects of petroleum hydrocarbons on algal physiology: review and progress report. In A.P.M. Lockwood, ed., *Effects of pollutants on aquatic organisms.* Cambridge University Press, New York, NY, pp. 107-125.

Walker, J.D. and R.R. Colwell. 1976. Long-chain *n*-alkanes occurring during microbial degradation of petroleum. *Can. J. Microbiol.* **22**:886-891.

Walker, J.D., R.R. Colwell and L. Petrakis. 1975. Degradation of petroleum by the alga, *Prototheca zopfii. Appl. Microbiol.* **30**:79-81.

Walker, J.D., R.R. Colwell and L. Petrakis. 1976a. Biodegradation rates of components of petroleum. *Can. J. Microbiol.* **22**:1209-1213.

Walker, J.D., R.R. Colwell and L. Petrakis. 1976b. Comparison of the biodegradability of crude and fuel oils. *Can. J. Microbiol.* **22**:598-602.

Ward, D.M. and T.D. Brock. 1976. Environmental factors influencing the rate of hydrocarbon oxidation in temperate lakes. *Appl. Environ. Microbiol.* **31**:764-772.

Westendorf, R.G. 1986. Performance aspects of volatile organics analysis by purge and trap capillary column gas chromatography with flame ionization detectors. Tekmar Technical Papers, Tekmar Company, Cincinnati, OH.

Westlake, D.W.S., A. Jobson, R. Phillippe and F.D. Cook. 1974. Biodegradability and crude oil composition. *Can. J. Microbiol.* **20**:915-928.

Wilkes, F.G. 1977. Laboratory microcosms for use in determining pollutant stress. In O. Hutzinger and S. Safe, Eds., *Volume 1. Aquatic Pollutants: Transformations and Biological Effects.* Pergamon Press Ltd., New York, NY, pp. 309-321.

Wolfe, D.A. 1987. Interactions of spilled oil with suspended materials and sediments in aquatic systems. In K.L. Dickson, A.W. Maki, and W.A. Brungs, eds., *Fate and Effects of Sediment-Bound Chemicals in Aquatic Systems.* Pergamon Press, New York, NY, pp. 299-316.

Worsey, M.J. and P.A. Williams. 1975. Metabolism of toluene and xylenes by *pseudomonas (putida(arvilla)* mt-2: Evidence for a new function of the TOL plasmid. *J. Bacteriol.* **124**:7-13.

Wyndham, R.C. and J.W. Costerton. 1981. Heterotrophic potentials and hydrocarbon biodegradation potentials of sediment microorganisms within the Athabasca oil sands deposit. *Appl. Environ. Microbiol.* **41**:783-790.

Wyndham, R.C. and J.W. Costerton. 1981. In vitro microbial degradation of bituminous hydrocarbons and in situ colonization of bitumin surfaces within the Athabasca oil sands deposit. *Appl. Environ. Microbiol.* **41**:791-800.

Zar, J.H. 1984. *Biostatistical Analysis, Second Edition.* Prentice-Hall, Inc., Englewood Cliffs, NJ. 718pp.

Zeyer, J., P.R. Lehrbach and K.N. Timmis. 1985. Use of cloned genes of *Pseudomonas* TOL plasmid to effect biotransformation of benzoates to *cis*-dihydrodiols and catechols by *Escherichia coli* cells. *Appl. Environ. Microbiol.* 50:1409-1413.

Zobell, C.E. 1946. Action of microorganisms on hydrocarbons. *Bacteriol Rev.* 10:1-49.

Zobell, C.E. 1950. Assimilation of hydrocarbons by microorganisms. *Adv. Enzymol.* 10:443-486.

Zobell, C.E. 1971. Sources and biodegradation of carcinogenic hydrocarbons. In *Proceedings of the Joint Conference on Prevention and Control of Oil Spills*, API, EPA, and U.S. Coast Guard, June 15-17, 1971, Washington, D.C. pp. 441-451.

APPENDIX A

Microbial species known to be responsible for the oxidative degradation of petroleum hydrocarbons.

Hydrocarbons	Microorganisms	Reference
Alkanes	*Pseudomonas* sp.	Gibson et al, 1974
	Pseudomonas oleovorans	Riser-Roberts, 1992
	Norcardia sp.	Gibson et al, 1974
	Pseudomonas aeruginosa	Van Eyk and Bartels, 1968
	Pseudomonas putida	Worsey and Williams, 1975
Propane	*Mycobacterium smegmatis*	Riser-Roberts, 1992
Butane	*Mycobacterium smegmatis*	Riser-Roberts, 1992
Pentane	*Mycobacterium smegmatis*	Riser-Roberts, 1992
Hexane	*Mycobacterium smegmatis*	Riser-Roberts, 1992
Decane	*Cornebacterium* sp.	Kester and Foster, 1963
Dodecane	*Cornebacterium* sp.	Hou, 1982
Tridecane	*Cornebacterium* sp.	Kester and Foster, 1963
	Pseudomonas aeruginosa	Van Eyk and Bartels, 1968
Tetradecane	*Micrococcus cerificans*	Riser-Roberts, 1992
	Cornebacterium sp.	Kester and Foster, 1963
Pentadecane	*Cornebacterium* sp.	Kester and Foster, 1963
Hexadecane	*Cornebacterium* sp.	Kester and Foster, 1963
Branched Alkanes	*Brevibacterium erythrogenes*	Riser-Roberts, 1992
	Cornebacterium sp.	Riser-Roberts, 1992
Alkenes	*Pseudomonas oleovorans*	Riser-Roberts, 1992
Aromatics	*Achromobacter* sp.	Rochkind et al, 1986
	Comamonas testosteroni	Kampfer et al, 1991
	Norcardia sp.	Gibson et al, 1974
	Pseudomonas sp.	Rochkind et al, 1986
	Pseudomonas aeruginosa	Rochkind et al, 1986
	Pseudomonas cepacia	Kampfer et al, 1991
	Pseudomonas fluorescens	Kampfer et al, 1991
	Pseudomonas putida	Gibson et al, 1974
	Cornebacterium sp.	Fedorak and Westlake, 1981
	Vibrio sp.	Fedorak and Westlake, 1981
	Acinetobaacter	Fedorak and Westlake, 1981
	Brevibacterium	Fedorak and Westlake, 1981
	Flavobacterium	Fedorak and Westlake, 1981
	Candida	Riser-Roberts, 1992
	Micrococcus	Fedorak and Westlake, 1981
	Alcaligenes sp.	Fedorak and Westlake, 1981
Benzene	*Moraxella*	Reineke et al, 1984
Toluene	*Escherichia coli*	Zeyer et al, 1985
Xylenes	*Nocardia* sp.	Gibson et al, 1974

Appendix B

Summary sheet of the Tekmar LSC 2000 Purge and Trap Concentrator and the Hewlett Packard 5890A Gas Chromatograph columns and analytical conditions.

Tekmar LSC 2000 Purge and Trap column and conditions:

Sample size: 5 ml
Valve, mount and line initial temperature: 30°C
Purge pressure: 140 kPa
Purge: 11 minutes at 42.6 cm/sec with N_2
Dry purge time: 4 minutes
Trap: Tenax/Silica Gel, 1/8" x 12", SS
Desorb preheat temperature: 175°C
Desorb temperature and time: 180°C for 4 min
Bake temperature and time: 180°C for 5 min

Hewlett Packard 5890A Gas Chromatograph column and conditions:

Column: SPB-5, 30 m x 0.53 mm ID, 1.5 µm film
Column head pressure: 30 kPa
Carrier Gas: Nitrogen
Nitrogen flow rate: 46.1 cm/sec
Hydrogen flow rate: 40 cm/sec
Air flow rate: 350 cm/sec
Column temperature program: 35°C/2 min//12°C/min to 225°C/5 min
Detector: Flame Ionization Detector
Integrator: Spectra-Physics 4290

Appendix C

Purge and Trap/Gas Chromatograph determinations of the area under the curve, for each of the listed certified hydrocarbon standards used to calculate hydrocarbon component concentrations in the JET-A MFC and JP-8 SAM microcosms.

Hydrocarbon Standards	Peak Area Under the Curve	Log Area Under the Curve	Concentration (ug/L)	Retention Time Minutes
2,4-Dimethylpentane	3114580	6.4934	10.8	2.51
2-Methylpentane	3273419	6.5150	10.8	1.85
2-Methylpropane	1110942	6.0457	3.0	1.33
3-Methylpentane	452626	5.6557	1.5	1.95
Benzene	11883196	7.0749	80.0	3.05
Butane	5425525	6.7344	9.0	1.43
Butylbenzene	2086230	6.3194	6.4	10.22
Cyclooctane	14266055	7.1543	32.0	8.24
Cyclopentane	845774	5.9273	15.2	1.63
Decane	1401309	6.1465	6.4	9.29
Dodecane	780875	5.8926	6.4	12.25
Ethylbenzene	392718	5.5941	1.5	6.90
Hexane	3874046	6.5882	10.8	2.13
Octane	3665145	6.5641	10.8	5.72
Pentane	280056	5.4472	19.4	1.64
cis-2-Pentene	1853569	6.2680	12.3	1.47
trans-2-Pentene	2250274	6.3522	14.9	1.50
Propane	31580	4.4994	3.0	1.18
Propylbenzene	3759580	6.5751	8.6	8.58
Tetradecane	406501	5.6091	4.4	14.82
Toluene	9977637	6.9990	21.6	5.05
Tridecane	432135	5.6356	4.4	13.58
m,p-Xylene	1095353	6.0396	1.8	7.06
o-Xylene	689156	5.8383	0.9	7.48

Appendix D

Log transformed degradation rate regressions for each hydrocarbon component in the JET-A MFC microcosm.

JET-A MFC MICROCOSM

WSF Treatment	Benzene	Toluene	*m,p*-Xylene
1%	y = 3.85 - 0.0139 R(sq) = 1.00	y = 5.00 - 0.0148 R(sq) = 1.00	y = 5.19 - 0.0090 R(sq) = 1.00
5%	y = 4.88 - 0.0141 R(sq) = 1.00	y = 6.26 - 0.0244 R(sq) = 0.93	y = 6.18 - 0.0159 R(sq) = 0.93
15%	y = 5.49 - 0.0151 R(sq) = 0.99	y = 6.81 - 0.0169 R(sq) = 0.97	y = 6.81 - 0.0151 R(sq) = 0.98
Mean 0% + 15%	y = 5.56 - 0.0137 R(sq) = 0.97	y = 7.07 - 0.0187 R(sq) = 0.94	y = 6.95 - 0.0136 R(sq) = 0.97
Mean 15% + 15%	y = 5.51 - 0.0106 R(sq) = 0.98	y = 6.98 - 0.0148 R(sq) = 0.93	y = 6.94 - 0.0121 R(sq) = 0.92

	o-Xylene	Ethylbenzene	Propylbenzene
1%	y = 4.96 - 0.0102 R(sq) = 1.00	y = 4.37 - 0.0113 R(sq) = 1.00	
5%	y = 5.94 - 0.0153 R(sq) = 0.94	y = 5.50 - 0.0191 R(sq) = 0.92	y = 4.52 - 0.0132 R(sq) = 1.00
15%	y = 6.53 - 0.0137 R(sq) = 0.98	y = 6.15 - 0.0177 R(sq) = 0.97	y = 5.33 - 0.0139 R(sq) = 1.00
Mean 0% + 15%	y = 6.68 - 0.0128 R(sq) = 0.94	y = 6.30 - 0.0157 R(sq) = 0.96	y = 5.33 - 0.0083 R(sq) = 0.92
Mean 15% + 15%	y = 6.60 - 0.0102 R(sq) = 0.95	y = 6.27 - 0.0145 R(sq) = 0.95	y = 5.37 - 0.0096 R(sq) = 0.91

	Butylbenzene	Cyclooctane	Decane
1%			y = 4.73 - 0.0157 R(sq) = 0.98
5%	y = 4.56 - 0.0126 R(sq) = 0.99	y = 4.53 - 0.0103 R(sq) = 1.00	y = 5.65 - 0.0142 R(sq) = 0.99
15%	y = 5.50 - 0.0149 R(sq) = 0.91	y = 5.33 - 0.0161 R(sq) = 0.98	y = 6.55 - 0.0224 R(sq) = 0.81
Mean 0% + 15%	y = 5.53 - 0.0074 R(sq) = 0.92	y = 5.56 - 0.0133 R(sq) = 0.91	y = 6.45 - 0.0120 R(sq) = 0.96
Mean 15% + 15%	y = 5.62 - 0.0080 R(sq) = 0.88	y = 5.57 - 0.0124 R(sq) = 0.88	y = 6.42 - 0.0106 R(sq) = 0.94

	Dodecane	Tridecane	Tetradecane
1%			
5%	y = 5.55 - 0.0182 R(sq) = 1.00		
15%	y = 6.28 - 0.0105 R(sq) = 0.95	y = 4.55 - 0.0152 R(sq) = 1.00	
Mean 0% + 15%	y = 6.75 - 0.0151 R(sq) = 0.94	y = 4.99 - 0.0161 R(sq) = 0.90	y = 4.51 - 0.0213 R(sq) = 0.42
Mean 15% + 15%	y = 6.74 - 0.0139 R(sq) = 0.88	y = 5.01 - 0.0108 R(sq) = 0.74	y = 4.63 - 0.0108 R(sq) = 0.90

Appendix E

Log transformed degradation rate regressions for each hydrocarbon component in the JP-8 SAM microcosm.

JP-8 SAM MICROCOSM

WSF Treatment	Benzene	Toluene	*m,p*-Xylene
1%			
5%	y = 4.22 - 0.0067 R(sq) = 0.42	y = 5.28 - 0.0116 R(sq) = 0.97	y = 5.71 - 0.0106 R(sq) = 0.96
15%	y = 4.72 - 0.0073 R(sq) = 0.93	y = 5.94 - 0.0088 R(sq) = 0.96	y = 6.42 - 0.0082 R(sq) = 0.97
Mean 0% + 15%	y = 4.66 - 0.0126 R(sq) = 0.98	y = 6.03 - 0.0170 R(sq) = 0.99	y = 6.20 - 0.0131 R(sq) = 0.94
Mean 15% + 15%	y = 4.69 - 0.0132 R(sq) = 0.98	y = 6.07 - 0.0180 R(sq) = 0.98	y = 6.40 - 0.0162 R(sq) = 0.98

	o-Xylene	Ethylbenzene	Propylbenzene
1%		y = 4.34 - 0.0022 R(sq) = 1.00	
5%	y = 5.47 - 0.0080 R(sq) = 0.95	y = 4.73 - 0.0102 R(sq) = 0.95	y = 5.27 - 0.0208 R(sq) = 1.00
15%	y = 6.14 - 0.0068 R(sq) = 0.97	y = 5.65 - 0.0099 R(sq) = 0.91	y = 6.17 - 0.0130 R(sq) = 0.92
Mean 0% + 15%	y = 6.01 - 0.0093 R(sq) = 0.96	y = 5.54 - 0.0126 R(sq) = 0.99	y = 6.17 - 0.0250 R(sq) = 0.90
Mean 15% + 15%	y = 6.07 - 0.0102 R(sq) = 0.99	y = 5.57 - 0.0143 R(sq) = 0.99	y = 6.13 - 0.0217 R(sq) = 0.89

	Butylbenzene	Cyclooctane	Decane
1%			
5%	y = 4.94 - 0.0102 R(sq) = 0.72	y = 4.36 - 0.0153 R(sq) = 0.98	y = 5.48 - 0.0051 R(sq) = 0.93
15%	y = 5.94 - 0.0125 R(sq) = 0.87	y = 5.33 - 0.0120 R(sq) = 0.92	y = 6.16 - 0.0053 R(sq) = 0.91
Mean 0% + 15%	y = 5.74 - 0.0107 R(sq) = 0.93	y = 4.96 - 0.0141 R(sq) = 0.96	y = 6.06 - 0.0119 R(sq) = 0.94
Mean 15% + 15%	y = 5.75 - 0.0113 R(sq) = 0.95	y = 4.99 - 0.0147 R(sq) = 0.95	y = 6.10 - 0.0125 R(sq) = 0.96

	Dodecane	Tridecane	Tetradecane
1%			
5%	y = 5.35 - 0.0054 R(sq) = 0.87	y = 5.18 - 0.0296 R(sq) = 0.80	y = 5.20 - 0.0351 R(sq) = 0.33
15%	y = 6.07 - 0.0047 R(sq) = 0.81	y = 5.66 - 0.0077 R(sq) = 0.74	y = 5.77 - 0.0251 R(sq) = 0.87
Mean 0% + 15%	y = 6.14 - 0.0069 R(sq) = 0.88	y = 6.11 - 0.0186 R(sq) = 0.81	y = 6.17 - 0.0212 R(sq) = 0.90
Mean 15% + 15%	y = 6.22 - 0.0093 R(sq) = 0.98	y = 6.26 - 0.0216 R(sq) = 0.89	y = 6.36 - 0.0268 R(sq) = 0.71

Appendix F

Calculated t values for significant differences between the hydrocarbon degradation slopes in the JET-A MFC and the JP-8 SAM 15%, 0%+15% and the 15%+15% WSF treatments when compared within their respective microcosm experiments.

	JET-A MFC (WSF)		JP-8 SAM (WSF)	
Hydrocarbon (WSF)	0%+15%	15%+15%	0%+15%	15%+15%
Decane [15%]	4.68	4.77	9.46	9.07
Decane [0% + 15%]				
Dodecane [15%]			2.91	7.32
Dodecane [0% + 15%]				4.11
Tridecane [15%]				
Tridecane [0% + 15%]				
Tetradecane [15%]				
Tetradecane [0% + 15%]				
Benzene [15%]	2.41	5.66	4.87	5.36
Benzene [0% + 15%]		2.55		
Toluene [15%]			8.37	8.13
Toluene [0% + 15%]		2.31		
m,p-xylene [15%]			3.80	6.58
m,p-xylene [0% + 15%]				3.48
o-xylene [15%]		2.84	4.21	7.78
o-xylene [0% + 15%]		2.52		2.47
Butylbenzene [15%]	3.73	3.50		
Butylbenzene [0% + 15%]				
Cyclooctane [15%]				
Cyclooctane [0% + 15%]				
Ethylbenzene [15%]			3.10	4.59
Ethylbenzene [0% + 15%]				7.10
Propylbenzene [15%]	3.21			5.02
Propylbenzene [0% + 15%]				

$t_{0.05,\,(2)\,28}$ = 2.056 (Zar, 1984).

Appendix G

Calculated t values for significant differences between the hydrocarbon degradation slopes in the JET-A MFC and the JP-8 SAM 15%, 0%+15% and the 15%+15% WSF treatments when compared between the microcosm experiments.

	JP-8 SAM		
JET-A MFC (WSF)	15% WSF	0%+15% WSF	15%+15% WSF
Decane [15%]	11.62	6.13	4.42
Decane [0% + 15%]	7.89		
Decane [15% + 15%]	5.50		
Dodecane [15%]	3.70	2.62	
Dodecane [0% + 15%]	5.52	4.46	2.88
Dodecane [15% + 15%]	5.18	4.13	2.52
Tetradecane [15%]			
Tetradecane [0% + 15%]		3.51	
Tetradecane [15% + 15%]			
Benzene [15%]	5.36	5.01	3.65
Benzene [0% + 15%]	5.35		
Benzene [15% + 15%]	4.29	2.50	3.17
Toluene [15%]	9.29		
Toluene [0% + 15%]	4.53		
Toluene [15% + 15%]	3.20	2.25	
m,p-xylene [15%]	6.05		
m,p-xylene [0% + 15%]	3.92		3.67
m,p-xylene [15% + 15%]	2.47		4.36
o-xylene [15%]	6.19	3.54	4.62
o-xylene [0% + 15%]	6.17	4.04	3.05
o-xylene [15% + 15%]	4.66		
Butylbenzene [15%]		2.21	2.13
Butylbenzene [0% + 15%]		2.97	3.06
Butylbenzene [15% + 15%]	2.32	3.65	3.91
Ethylbenzene [15%]	4.47	5.70	5.63
Ethylbenzene [0% + 15%]	3.17		
Ethylbenzene [15% + 15%]			
Propylbenzene [15%]			
Propylbenzene [0% + 15%]			4.57
Propylbenzene [15% + 15%]			3.49

$t_{0.05,\,(2)\,26} = 2.056$ (Zar, 1984)

INVESTIGATION OF THE EFFECTS OF A PULSED RELEASE OF JET-A TURBINE FUEL FROM SEDIMENTS USING A MODIFIED MIXED FLASK CULTURE (MFC) MICROCOSM

by

Randy S. Sandberg

INVESTIGATION OF THE EFFECTS OF A PULSED RELEASE OF JET-A TURBINE FUEL FROM SEDIMENTS USING A MODIFIED MIXED FLASK CULTURE (MFC) MICROCOSM

by

Randy S. Sandberg

Accepted in Partial Completion

of the Requirements for the Degree

Master of Science

Dean of Graduate School

Advisory Committee

Chair

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this thesis is allowable only for scholarly purposes. It is understood, however, that any copying or publication of this thesis for commercial purposes, or for financial gain, shall not be allowed without my written permission.

Signature Randy L Sandberg

Date 11-11-93

Bellingham, Washington 98225 (206) 650-3000

INVESTIGATION OF THE EFFECTS OF A PULSED RELEASE OF JET-A TURBINE FUEL FROM SEDIMENTS USING A MODIFIED MIXED FLASK CULTURE (MFC) MICROCOSM

A Thesis Presented To

The Faculty Of

Western Washington University

In Partial Fulfillment

Of The Requirements for the Degree

Master of Science

---

by

Randy S. Sandberg

November 1993

ABSTRACT

The aquatic toxicity information used to satisfy regulatory requirements under FIFRA are generated under a tiered testing sequence with nearly all decisions regarding registration based on the results of single species tests. Over the last 15 years, a variety of multispecies aquatic toxicity tests have been developed with the hope that the increased complexity of the test system would result in a more realistic, community-level response to contamination. Sediments are oftentimes a major repository for contaminants introduced into surface waters. The science of sediment toxicology itself, however, has been described as being in its infancy due to the failure to incorporate ecosystem disturbance into toxicity assessments.

This study investigates both the methods and the ecosystem level effects of producing a simulated release of a complex hydrocarbon mixture from sediments using a 60-day one l modified Mixed Flask Culture (MFC) microcosm. Treatment sediment groups consisting of six microcosm replicates were spiked with 0, 2, 10 and 25 microliters of Jet-A based on the results of preliminary acute 10-day freshwater sediment amphipod bioassays using *Hyalella azteca* as the test species. For each test chamber, a spiked layer of Standardized Aquatic Microcosm (SAM) sediment was encapsulated under an overlying layer of coadapted MFC silica sand and detritus. Data were examined using both conventional univariate, as well as newly developed multivariate techniques.

Analysis of The Jet-A using purge and trap gas chromatography revealed a slow pulsed release of the test material from the spiked layer. Univariate results

of the functional parameters indicated that an initial period of perturbation occurred followed by a stable state.Effects were apparently caused by the transfer perturbation of the spiking procedure, as well as the effects of thehydrocarbon mixture. Univariate results of structural parameters indicated that treatment effects generally detectable through the entire test, that a general initial imbalance in population sizes existed at the beginning of the treatment period on day zero, and that no apparent stability of the control group or recovery of the system from perturbation was apparent. Virtually all multivariate techniques were able to distinguish statistically significant responses of the system to treatment despite the relatively small proportion of Jet-A used in the test.

These results suggested that although both the cross-inoculation procedure and the spatial scale of the MFC system may be inadequate, the method of incorporating spiked sediment into the MFC is a useful technique and may merit further study. The observed instability of the reference group and the failure of the system to return to a pre-exposure state were also not incompatible with the observations of other questioning the existence of stability in natural systems.

ACKNOWLEDGEMENTS

I am indebted to Wayne G. Landis and April Markiewicz of the Institute of Environmental Toxicology and Chemistry (IETC), Huxley College, Western Washington University for their advice, patience, and encouragement in all phases of this project. I also wish to thank Tim Hall and Jack Hardy for their advice and support. I would also like to thank Lisa D.Holmquist and the rest of the IETC staff, as well as the staff of Huxley College, for their support.

Appreciation is also expressed to Emily Peele for her advice and the use of her epifluorescent microscope, to Nancy Elkins for her stockroom support, and Richard Mayer for the use of his gas chromatograph.

I am especially grateful to my wife, Karen, and my daughter Natalie, for their financial, moral, and loving support during this period.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Regulation for the testing of the environmental fate and effects of chemicals produced in the United States is conducted under the Federal Water Pollution Control Act, the Toxic Substances Control Act, and the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA). FIFRA is unique from other legislation in that it (1) regulates known toxic chemicals that are manufactured for direct environmental application, and (2) the method, amount and timing of entry into the environment can be predicted and regulated (Bedford, 1984). Under FIFRA, manufacturers are required to provide registration information regarding environmental fate, actual field dissipation, and aquatic toxicity information using laboratory acute and chronic tests, simulated field tests, and full field studies on nontarget species (Bedford, 1984; Urban and Cook, 1986).

The aquatic toxicity information used to satisfy these requirements are generated under a tiered testing sequence (Urban and Cook, 1986). Acute toxicity information for both fish and invertebrates is required. Other testing, including chronic tests, simulated field tests and full field tests, are often not required based on the results of the acute testing. Recently, the United States Environmental Protection Agency has suspended the requirements for conducting ecosystem level studies for pesticide registration based on the limited information derived from these tests, as well as the high costs in both effort and expense (Fisher, 1992). Even before this new regulatory decision, few chemicals have ever undergone a full tiered testing and nearly all decisions regarding pesticide registration is based on information derived from single-species tests.

Many researchers have criticized this approach due to the fact that single species tests may not be adequate predictors of potential effects on communities

and ecosystems (Cairns, 1986; Kimball and Levin, 1985). Acute lethality testing was originally designed for the bioassay of drugs used in individual organisms that were not amenable to chemical study and is ideal for that purpose (Moriarty, 1988). The use of the acute LD_{50} over the last 40 years to predict the toxicity of xenobiotics is much more uncertain. The value will vary with species, strain, age, environmental conditions and genomic structure, and does not take into account sublethal effects. Clearly, the potential rate of increase for a population, r, will decrease not only with an increase in the death rate predicted by the LD_{50}, but also a shrinking birth rate through interferences in fertility and fecundity, as well as developmental effects.

Multispecies Tests

Ecosystem Properties

Since all naturally occurring organisms live in ecosystems, the structural and functional properties of ecosystems determine the context in which populations and communities of these organisms develop, persist, and interact. Ecosystems can be viewed as simply energy processing units, converting incoming solar energy into chemical energy and finally heat. Because this process requires a supply of inorganic nutrients, a certain portion of energy is utilized in obtaining and recycling these nutrients.

Thus, ecosystem function can be viewed as a combination of energy and matter flow. The production, or anabolic process, is simply the conversion of inorganic matter to organic matter utilizing the sun's energy in photosynthesis. The catabolic, or regenerative process, is simply the release of stored organic chemical energy as heat and the elements are returned to inorganic form.

Xenobiotic effects on ecosystems have the potential to disturb all of the components of the ecosystem either directly or indirectly. The immediate

effects of xenobiotics released into the environment are on individual organisms, either through direct toxicity or by altering the environment (Moriarty, 1988). However, the ecological significance resides in the indirect impact on populations of species. For example, a high mortality of a particular species population can have little or no ecological significance. On the other hand, a xenobiotic that kills no individuals of a population but retards development can have a severe ecological impact. Thus, xenobiotics with no direct toxicity can have severe ecological impacts.

Theoretically, a reduction in population size of a species due to the introduction of xenobiotics could potentially increase the sizes of other populations of aesthetically, ecologically, or economically important species due to the alteration of interspecies relationships in early life stages, resulting in the survival of more individuals to maturity. Effects such as this simply could not be determined through single-species testing of indigenous or surrogate species and could potentially be evaluated in a multispecies test system.

Microcosms

Over the last 15 years a variety of multispecies aquatic toxicity tests have been developed with the hope that the increased complexity of the test system would result in a more realistic, community-level response to xenobiotics (Giddings, 1981; Leffler, 1984; Taub and Read, 1982; Touart, 1988). These multispecies tests have been recommended as logical and meaningful intermediates between traditional population-based single-species tests and uncontrolled natural ecosystems for evaluating the effects of xenobiotics (Cairns, 1984; Giddings, 1981; Kimball and Levin, 1985). The size of multispecies tests can range from 1 l microcosms, as in the Mixed Flask Culture (MFC) (Leffler, 1984), to the thousands of liters commonly used in pond mesocosms for pesticide registration testing (Touart, 1988). Although no clear size distinction

has been made between microcosms and mesocosms, microcosms are generally regarded as 1-4 l static, open, freshwater systems (Leffler, 1984).

Realism vs. Generality

A common distinction is often made in the literature between "generic" microcosm systems, which mimic no specific system in any detail but exhibit properties common to all systems, and systems that simulate some specific ecosystem in lesser or greater detail (Giddings, 1981). A realistic system designed to simulate a particular natural ecosystem does so at the expense of generality (Giddings, 1981). Conversely, a generic system designed to simulate broad categories of natural systems does so at the expense of simulating any one system. As stated by Stay et al. (1989a): In general, microcosms proposed for screening toxic chemicals are generic systems that do not replicate any specific natural system but simulate properties common to many systems. Generic microcosms are of two types: (1) those derived from mixing stock monocultures (i.e. the Standardized Aquatic Microcosm (SAM), Taub and Read, 1982), and (2) those derived from natural ecosystems (i.e., the Leffler microcosm (MFC), Leffler, 1984).

The Standardized Aquatic Microcosm (SAM) system (ASTM E 1366-91) is a 60-day three l test system defined as to species, media, and substrate. Replicability and repeatability are theoretically obtained by inoculating specific quantities of traditionally laboratory cultured test organisms generally used in single species tests into containers of artificially prepared chemically defined sterile test media and substrate. Ten species of algae are added on day zero, six species of protozoa and animals on day four, and the test material on day seven. The system is then kept on a 12/12 hour light/dark schedule at a specific temperature and is monitored twice weekly for structural and functional properties. Evaluation is usually based on the results of graphical

representations of significant differences between treatment groups based on ANOVA for each measured parameter.

The Mixed Flask Culture (Leffler, 1984), on the other hand, is a multispecies test system utilizing organisms derived from natural inocula. This protocol is based on the assumption that ecosystem-level functional properties of naturally derived systems are independent of structure and microcosms derived from a variety of sources will respond consistently to the same xenobiotic based on these functional properties. Also, since the microcosms are derived from natural sources, extrapolation to natural ecosystems is considered to be more accurate although this has not been substantiated.

In this system, an initial 2 l inoculum for a 40 l stock community is obtained from natural sources and allowed to mature for three months in a laboratory. This period allows the stock culture to theoretically coadapt into a stable species assemblage. One l beaker microcosms containing a small amount of silica sand sediment and 950 ml of T82MV (ASTM E 1366-91) media are then inoculated with 50 ml of this stock community. These are then allowed to "mature" for six weeks in an incubator under similar conditions to the SAM. Twice weekly cross inoculations, additions of sterile media, and weekly rotations are performed prior to day zero to theoretically provide for consistency among replicates, simulate immigration, and permit reintroduction of extirpated species. Both functional and structural variables are then monitored on a schedule of decreasing frequency with emphasis being placed on functional variables. Structural components are not monitored with the same degree of detail as in the SAM.

Structure vs. Function

Increasingly however, evidence suggests that aquatic ecosystem structural changes are apparent before any functional process changes become

detectable (Cairns and Pratt, 1986; Odum, 1990; Odum, 1985; Pratt, 1990; Schindler, 1987). Functional parameters are robust and are primarily substrate driven and limited, and are not overly influenced by the particular biological machinery processing these substrates. Consequently, where anthropogenic stressors do not affect the supply of these substrates, functional variables show little effect. As an example of this, Stay et al. (1988), in analyzing the effects of Fluorene on MFC microcosms developed from four natural communities, reported that slight but significant changes in functional variables gave no indication of the almost complete elimination of some zooplankton populations, suggesting further that effects of function give little insight into structure.

Community structure data however, although high in information content, is usually difficult to analyze and interpret. Individual species, being in a natural setting, are usually clumped in negative binomial distributions and may not even be present in some replicates. Combined with the impacts of the xenobiotics themselves, this can lead to invalidations of the assumptions of normality and homogeneity of variance required by ANOVA, thereby reducing the statistical power of the procedure. These problems have resulted in some cases, in the selection and use of community parameters for the interpretation of tests results based on their statistical characteristics, in seeming disregard of their ecological relevance. These have primarily been functional variables due to the more normal distributions of measurements even though function gives little insight into structure. However, ANOVA has been shown to be robust despite violations of assumptions as long as there are equal numbers of replicates and has been used successfully (Zar, 1984), although problems with type II errors remain (failing to reject a false null hypothesis due to interferences from variance).

The Genome

An aspect of multispecies testing not often encountered in the toxicological literature involves discussions of the influence of genomic structure on the results of multispecies toxicity tests. Individual species populations can not only be regarded as collections of individuals, but as collections of genes (Moriarty, 1988). Xenobiotic stress, acting selectively through direct or indirect effects on individuals in a population, can alter the composition and size of that particular populations genome. This happens, in large part, in a stochastic fashion.

Individuals within that population may, or may not have the particular genetic makeup neccessary to effectively cope with the particular direct or indirect effects of that xenobiotic. This may result in selection against individuals in that population and a subsequent restriction of the gene pool due to the elimination of all of that particular organisms genes. This sets the stage for subsequent random genetic drift (founder effect) and island biogeography to occur within the test unit.

These effects are similar to both those occuring due to natural stressors on a population, and to the partitioning of natural systems in small enclosures. Enclosing small portions of ecosystems is identical to the establishment of a new sub population in nature and results in initial differences in gene frequencies from the original populations. In addition to this, larger proportions of the total gene pool are attributed to each individual organism and their particular genetic combination. This allows more freedom for stochastic events to influence gene frequencies, and can produce a treatment effect similar to that produced by xenobiotics and natural stressors.

Criticisms

Despite the strong and convincing theoretical ecological arguments that can be made endorsing the use of multispecies tests (Cairns, 1983; McMahon et al.,

1978), several concerns remain from both a regulatory as well as a scientific standpoint. These can be generally condensed into six main points (Cairns 1993):

1. Are the results more predictively accurate than single species tests? (Mount, 1987)
2. Are multispecies results more easily extrapolated to natural systems? (Heath, 1980)
3. How replicable and reproducible are the results?
4. Will the magnitude of natural variation prevent the detection of effects?
5. Are suitable end points possible to determine?
6. Are multispecies tests cost effective?

The first five points illustrated here are instructive in that they illustrate the current limits of the field of ecotoxicology. As Cairns (1993) points out, it is gradually being realized that on both spatial and temporal scales, most studies of ecotoxicology are inadequate. Population ecology, rather than ecosystem and landscape ecology, has been the predominant area of interest in the past. Consequently, conceptual and statistical difficulties arise when attempts are made to evaluate and extrapolate the results of multispecies tests to natural systems.

Multivariate Analysis

In view of these problems, a major difficulty in the evaluation of multispecies tests has been in analyzing results on a level consistent with the goals of the toxicity test (Landis et al., 1993a; 1993b; 1993c). Detectable changes in

individual population dynamics within the test system must be incorporated into a community level response to the xenobiotic in question.

Conventional tests of significance, such as Analysis of Variance (ANOVA), have been successfully used to examine significant differences of single variables between treatment groups. However, due to the temporal dependence of the data, an increasing likelihood of making a type II error (accepting a false null hypothesis), and difficulties in representing the data set graphically, there are problems associated with these tests. ANOVA's calculated within a sampling day for each variable, besides being difficult to depict graphically, may give misleading results. Problems with increasing within group variation over time decreases the probability detecting effects and increases the probability of making type II errors. Questions of when to reject the null hypothesis when significant differences are found for one or a few variables also arise when different numbers of variables are monitored for each test.

Conquest and Taub (1989) have developed a method that overcomes several of these problems. Called the Interval of Nonsignificant Difference (IND), this method is easily able to graphically depict significant differences from the control mean based on ANOVA within sampling days. By depicting these intervals over time, this method corrects for the likelihood of making type II errors. Graphical depictions of the magnitude of difference from the non treated group required to obtain significance, as well as which treatment groups are significantly different from the non treated group, are possible. This method is routinely used in SAM experiments and is valid for use in other applications.

However, while this is a useful method for examining data on a variable by variable basis, it fails to incorporate these individual dynamics into a single community-level response. Multivariate methods have shown promise in evaluating all of the variables holistically, (Johnson, 1988a; Johnson, 1988b;

Kersting, 1988; Matthews et al., 1991a; Matthews et al., 1991b; Smith et al., 1990), thereby taking a step closer to evaluating and integrating responses on ecosystem and landscape levels rather than extrapolating individual population responses from single and multispecies tests to natural complex ecosystems.

Sediments

Sediments are oftentimes a major repository for contaminants introduced into surface waters (Lyman, 1984). Since petroleum hydrocarbons, other organics, and heavy metals tend to sorb to it, sediment often accumulates contaminant concentrations several orders of magnitude higher than that found in the water column (Lee and Jones, 1984). There is known to be a continual flux of inorganic (Shaw and Prepas, 1990) and organic compounds (Meyer-Reil, 1987) through the sediment-water interface.

In addition to accumulating contaminants to a much higher degree than the water column, sediments are also orders of magnitude more permanent and serve as a better record of past contamination (Burton, 1991). This has led to increased monitoring of sediment contamination and benthic macroinvertebrate communities by regulatory agencies (Southerland et al., 1992; USEPA, 1987).

Sediment Toxicology

The science of sediment toxicology has been described as being in it's infancy due to the failure to incorporate ecosystem disturbance into toxicity assessments (Burton, 1991). During the late 1970's and early 1980's, it became apparent that physiochemical and biological relationships between sediment contaminants and the sediment environment were complex and variable and not easily manageable using chemical criteria (Lee and Jones, 1984). This led to increased regulatory interest and research activity into better methods for assessing contamination (USEPA, 1987). Consequently, numerous single-

species assays were developed for the assessment of sediment toxicity to a variety of organisms. Subsequent reviews of studies using these assays has revealed that each test species or selected endpoint was the most sensitive at one time or another (Burton, 1991) and that no species is the most sensitive to all xenobiotics (Sloof et al., 1986). Other studies have demonstrated the importance of using multiple assays of several species to evaluate toxicity due to this variation in sensitivity between species for each specific contaminant in each environmental setting (Pontasch et al., 1989, Burton et al., 1987; Wiederholm and Dave, 1989).

These surrogate responses are simply quantified on the basis of sample toxicity and the effects are extrapolated to *in situ* conditions. Although this satisfies the objectives of defining sample toxicity to the test species, they do little to document and define ecosystem toxicity (Burton, 1991). This disparity is becoming increasingly more obvious as examples are published in the literature (Pontasch et al., 1989). Since significant cases of acute toxicity have been encountered only infrequently (Chapman, 1989) and subacute levels of contamination with the potential to disrupt ecosystem structure and function more common, the need for the investigation of a test system providing an ecosystem level response to contaminated sediment is clearly indicated.

Turbine Fuels

Leaking underground storage tanks are a major source of groundwater contamination by complex mixtures of petroleum hydrocarbons containing hazardous compounds regulated by the EPA (Hutchins et al., 1991). There are approximately two million underground petroleum storage tanks in the U.S. and there have been 90,000 confirmed releases reported in the U.S. during 1988 and 1989 (OUST, 1990). Despite conscientious oil spill prevention programs,

petroleum hydrocarbons are occasionally released from pipeline systems (Guiney et al., 1987). The fate and effects of these materials has not been studied as thoroughly as in the marine environment despite the relatively high volumes of these contaminants in freshwater systems (Guiney et al., 1987; U.S. Coast Guard, 1982-83).

Because of their wide availability, the low operating cost associated with their use, and the reliability of the turbine power plants using them, turbine (jet) fuels are one of the primary internal combustion fuels available worldwide. Any spills of aviation fuel in the U.S. will likely involve one of the major formulations of turbine fuel: Jet-A, JP-4, JP-5, or JP-8 (Landis et al., 1993a; Landis et al., 1993c). Several such spills are well documented and have recently been the subject of other types of research. This work has mainly focused on the bioremediation potential for contaminated aquifers and soil (Aelion and Bradley, 1991; Hutchins et al., 1991; Madsen et al., 1991; Song et al., 1990).

Turbine fuels also offer advantages as model complex toxicants for toxicological research. Due to their use primarily as aviation fuel, turbine fuels are produced to more stringent standards than other types of internal combustion fuel (ASTM D 1655-89). The characteristic combustion/retention "hump" of individual organic components seen in gas chromatographic outputs of virtually all complex fuels, is required for the smooth and efficient running of most internal combustion engines and is more tightly controlled as to purity, chemical constituency and relative composition in turbine fuels for safety reasons. For example, automotive gasolines produced in the "blender" area of a refinery are typically "tested" by burning them in single cylinder test engines (ASTM D 2699-89, ASTM D 2700-89). The types and relative amounts of individual compounds used in their formulation can be manipulated by the refinery technicians until volatility requirements are met and the engines no

longer give the "pinging" sound characteristic of low quality fuel (ASTM D 439-89). Consequently, fuels can be manufactured to performance standards utilizing whatever the refinery has ample supplies of, resulting in quite different complex mixtures between manufacturers even within a given manufacturer on a given day. Turbine fuels, however, being produced to much more exacting standards for safety reasons, are much more likely to be very similar mixtures. Consequently, any multispecies test results and associated risk analysis are much more likely to be valid for a number of environmental applications.

Purpose

The purpose of this research was to evaluate both the methods and the ecosystem level effects of producing a simulated release of a complex hydrocarbon mixture from sediments using a 60-day one l modified Mixed Flask Culture (MFC) microcosm. Treatment sediment groups consisting of six microcosm replicates were spiked with 0, 2, 10 and 25 microliters of Jet-A based on the results of preliminary acute 10-day freshwater sediment amphipod bioassays using *Hyalella azteca* as the test species. For each test chamber, a spiked layer of Standardized Aquatic Microcosm (SAM) sediment was encapsulated under an overlying layer of coadapted MFC silica sand and detritus. Data were examined using both conventional univariate, as well as newly developed multivariate techniques.

Analysis of The Jet-A using purge and trap gas chromatography revealed a slow pulsed release of the test material from the spiked layer. Univariate results of the functional parameters indicated that an initial period of perturbation occurred followed by a stable state. Effects were apparently caused by the transfer perturbation of the spiking procedure, as well as the effects of the hydrocarbon mixture. Univariate results of structural parameters indicated that

treatment effects generally detectable through the entire test, that a general initial imbalance in population sizes existed at the beginning of the treatment period on day zero, and that no apparent stability of the control group or recovery of the system from perturbation was apparent. Virtually all multivariate techniques were able to distinguish statistically significant responses of the system to treatment despite the relatively small proportion of Jet-A used in the test.

These results suggested that although both the cross-inoculation procedure and the spatial scale of the MFC system may be inadequate, the method of incorporating spiked sediment into the MFC is a useful technique and may merit further study. The observed instability of the reference group and the failure of the system to return to a pre-exposure state were also not incompatible with the observations of other questioning the existence of stability in natural systems.

METHODS AND MATERIALS

Reagents

All chemicals used in either the culture of the laboratory organisms used in the study, the preliminary acute tests, or in the Mixed Flask Culture were reagent grade as specified in the ASTM method for the Standardized Aquatic Microcosm (SAM)(ASTM E 1366-91, 1991). Jet-A turbine fuel was obtained from Fliteline services in Bellingham. The shipment lot number was recorded and is on file[1].

Acute Tests

In order to determine appropriate concentrations of Jet-A to use in the MFC, both range-finding and definitive acute, 10-day amphipod bioassays using *Hyalella azteca* were conducted according to the ASTM protocol (ASTM E 1383-90). These tests were conducted using the same sterile sediment and media as used in the MFC. For a summary of test conditions, see Table 1.

For test chambers, 1 l borosilicate glass beakers (Pyrex no. 1000) that had been washed in a non phosphate laboratory detergent (Labtone), rinsed in distilled tap water, acetone rinsed, rinsed three times in distilled tap water, and dried for two hours at 105 ^{o}C were used. Each chamber was randomly assigned a number, treatment group, and shelf position in a temperature and light controlled room before being assembled.

Assembly included the addition of 100 ml of silica sand that had been acid washed and rinsed to pH seven with tap distilled water. Individual test chambers were then spiked with the appropriate amount of Jet-A added to the 100 ml of silica sand sediment according to treatment group. Each chamber's

[1]Institute of Environmental Toxicology and Chemistry (IETC), Huxley College, Western Washington University, Bellingham, WA 98225.

TABLE I

Summary of test conditions for conducting acute amphipod bioassays

Organisms	
Species	20 *Hyalella azteca*
Size	4 mm in length
Feeding	Every three days, 14 mg or more per unit
Experimental design	
Test type	static
Test vessel	1 l beakers, covered and aerated
Volume	1 l total
Replicates x concentrations	3x4 (range) 6x4 (definitive)
Concentrations	0 µls, 10 µls, 100 µls, and 1000 µls (range) 0 µls, 250 µls, 500 µls, and 750 µls (definitive)
Addition of toxicant	Day 0
Duration	10 days
Endpoint	Death
Physical and chemical parameters	
Temperature	20 °C +/- 1 °C
Photoperiod	12 hours light/12 hours dark
Medium	900 ml T82MV
Sediment	100 ml acid washed white quartz sand
Water quality measurements	
Day 0 and day 10	dissolved oxygen conductivity hardness pH alkalinity

sediment was first sprinkled with the appropriate amount of Jet-A from a Hamilton chromatography syringe. The sediment was then stirred with a sterile glass rod and the test chamber immediately covered with a 150 x 15 mm diameter petri dish to minimize evaporation of the Jet-A. The chamber was then held on a vortex vibrational mixer for 15 seconds to further homogenize the

sediments. Finally, to avoid mixing during filling, a 100 x 15 mm diameter sterile petri dish was placed in the chamber over the sediment using sterile forceps and the chamber was filled to one l with the sterile media used in the SAM (850 ml T82MV)(ASTM E 1366-91). The 100 x 15 mm sterile petri dish was then carefully removed with sterile forceps so as to minimize disturbance of the sediment-water interface. Twenty carefully acclimated and sized test organisms from the stock culture[2] were then added. Chambers were then individually covered with plexiglas covers, placed on their assigned shelf positions, and aerated with charcoal filtered air for ten days at 20 ^{o}C +/- 1 ^{o}C on a 12/12 hour light/dark schedule. The test organisms were fed 14 mg of Purina brand[3] rabbit pellets per test chamber every three days beginning on day zero (Ingersoll and Nelson, 1990).

Dissolved oxygen, pH, conductivity, hardness, and alkalinity were measured both at the beginning and end of the tests. On day zero, all chambers were measured for dissolved oxygen. For the acute range-finding test, pH and alkalinity were assumed to be equal to the values for the stock T82MV, which is recorded for all sterile T82MV media produced and is kept on file[4]. Two sterile 18 l carboys were required to obtain the appropriate volume for the definitive test, alkalinity and pH and were measured on a single aliquot of a 50/50 mixture of the two carboys. Similarly, conductivity and total hardness were measured on a single aliquot of the final T82MV stock solution used for each test.

On day 10, all chambers were measured for dissolved oxygen and pH in both tests. Conductivity, hardness, and alkalinity were measured on a single

[2]Institute of Environmental Toxicology and Chemistry (IETC), Huxley College, Western Washington University, Bellingham, WA 98225. Original stock culture obtained from: Eugene Green, National Fisheries Research Contaminant Center, 4200 New Haven Rd. Columbia, MO. 65201-9634. Received 12/6/91.
[3]Purchased at Hohl Feed and Seed, 1324 Railroad Ave, Bellingham, WA 98225.
[4]Institute of Environmental Toxicology and Chemistry (IETC), Huxley College, Western Washington University, Bellingham, WA 98225.

randomly chosen chamber from each treatment group. During the test, dissolved oxygen was measured daily on those chambers observed to be having difficulty with aeration to ensure that the minimum dissolved oxygen requirements of 40 % of saturation were met. For a summary of details of water quality methods, see Appendix I.

Data were then analyzed for normality and homogeneity, and subjected to an appropriate definitive statistical analysis under a null hypothesis of treatment having no effect. LC_{50}'s were then determined graphically.

Mixed Flask Culture

The Mixed Flask Culture (MFC) protocol has been previously described (Leffler, 1984). Fifty ml of acid washed white quartz sand sediment was added to 30 individual one l Pyrex beaker test chambers (Pyrex no. 1000). Each chamber had been washed in Labtone, rinsed in tap water, 2N HCL washed and then rinsed ten times in tap distilled water. Fifteen µg/l of $NaHCO_3$ and 900 ml of T82MV were then added to the test chambers before they were inoculated with 50 ml of a naturally derived stock community from a 40 l acid washed aquarium. This stock culture, containing 2-3 cm of similarly acid washed white quartz sand sediment and 38 liters of sterile T82MV media, had originally been inoculated with 2 liters of water from a variety of natural sources in the Bellingham (Washington) area. This culture was allowed to mature for three months prior to being used as an inoculum for the test chambers. Each beaker was then randomly numbered and placed on the bottom shelf of a 20 ^{o}C +/- 1 ^{o}C, 12/12 hour light/dark incubator.

Once weekly for six weeks following inoculation from the stock culture, the 1 l test chambers were cross inoculated to provide for consistency among replicates through simulation of immigration and permitting the reintroduction of extirpated

species. Cross inoculation consisted of stirring and removing 100 ml from each of the test chambers and adding this to a sterile 6 l erlenmeyer along with 300 ml of the original stock culture. This mixture was vigorously swirled and then 110 ml was redistributed to each of the test chambers. Each chamber was then topped up to 1 l with fresh T82MV. During this period, rotation within the incubator was carried out twice weekly.

At the end of the six week maturation period, the individual chambers were randomly selected and culled to 24 chambers based on preliminary predawn and late afternoon dissolved oxygen and pH readings. Since all chambers had the required 4 mg/l predawn and 11 mg/l late afternoon dissolved oxygen levels, microcosms having the highest deviations of dissolved oxygen from the mean values were selected. The selected microcosms were then examined for the required functional groups.

The individual chambers were then randomly split into four groups of six microcosms and spiked accordingly. Each of the remaining 24 chambers was randomly assigned a new number, treatment group, shelf, and shelf position in a 20 ^{o}C +/- 1 ^{o}C 12/12 hour light/dark incubator. Each of six shelves held one test chamber from each of four treatment groups. Each chamber was covered with a 150 x 15 mm sterile petri dish. Rotation within the incubator continued on a weekly basis until the end of the experiment. No reinoculations were performed after spiking.

Spiking

Each individual test chamber was sediment spiked with Jet-A according to treatment group. A new, identically cleaned and numbered 1 l chamber, containing an additional 50 ml of SAM sediment comprised of acid washed white quartz sand and both powdered cellulose and chitin in a 0.5gr(cellulose or chitin)/200 g sand ratio (ASTM 1399-91) was first injected with an appropriate

amount of Jet-A using a Hamilton chromatography syringe. The 0 µl group received only distilled deionized water while treatments 2, 3, and 4 received 2, 10 and 25 µl of Jet-A. A 150 x 15 mm diameter petri dish was then immediately placed over the chamber to minimize evaporation while being held on a vortex vibrational mixer for 15 seconds to homogenize the spiked sediments. Finally, to avoid mixing during transfer, a 100 x 15 mm diameter sterile petri dish was immediately placed in the chamber over the treated sediment using sterile forceps and the original microcosm was transferred over to the new dosed chamber by gently pouring and scraping with a sterile rubber policeman. The 100 x 15 mm dish was then gently removed using the sterile forceps to minimize disturbance of the underlying, spiked sediment.

Regular sampling was then carried out in accordance with the established SAM (ASTM E 1366-91) and MFC (Leffler, 1984) protocols. Sampling included dissolved oxygen, pH, turbidity, and organism numerical densities twice weekly on sampling days (Tues & Fri). Dissolved oxygen was monitored as in the SAM protocol in order to calculate P/R ratios. This included predawn and late afternoon measurements on days prior to sampling (mondays and thursdays) in order to obtain the necessary data to calculate the P/R ratio. pH and turbidity were measured just prior to disturbance in the morning on the day of sampling.

For numerical densities, each microcosm was vigorously stirred and the sides scraped with sterile rubber policemen specific for treatment group. Each chamber was then subsampled for algal, protozoan, and "large" organism (invertebrate) counts. Algae was counted using a palmer nanoplankton counting chamber and a Zeiss microscope at 400x. A total of 50 cells of each category were counted or 25 fields, whichever was reached first. Densities were then calculated utilizing the known media volumes of the Palmer cells and the area of the microscope fields. For protozoan counts, *Paramecium bursaria* and Rotifers

were counted as per ten 50 µl drops, ciliates and flagellates were counted as per two 10 µl drops dispensed with a calibrated automatic micropipettor. All were viewed with a dissecting scope. Large organisms were counted using a SAM sampling device. 300 ml were removed from each microcosm via sterile, treatment specific SAM samplers and mason jars. This volume was sequentially counted for large organisms in small increments before being placed back into the test chamber. For a summary of test conditions, see Table 2.

Gas Chromatography

Gas Chromatography samples were collected both the evening prior to sampling and the day of sampling of the MFC in order to track the pulsing in concentration in the dosed treatment groups as a result of the disturbance of sampling. To conserve volume, one sample was taken from each of the treatment groups on a rotational basis both the evening before sampling and then immediately after stirring vigorously the morning of sampling for each sampling day. Four ml of media were removed from the approximate center of each of the sampled chambers using a 10 ml disposable pipette and was stored at 4 ^{o}C in a cleaned and acid washed screw top test tube. These samples were then analyzed using purge and trap (P&T) gas chromatography. This was performed using a Tekmar LSC 2000 Purge and Trap (P&T) concentrator in tandem with a Hewlett Packard 5890A Gas Chromatograph and a Flame Ionization Detector (FID). Deionized distilled water blanks were used to verify the P&T and GC columns cleanliness prior to analysis of the sample. A 3.5 ml sample was injected into a 5 ml sparger, purged with pre-purified nitrogen gas for 11 min and dry purged for 4 min. Volatile hydrocarbons, purged from the sample and collected on the Tenax/Silica Gel column, were desorbed at 180 ^{o}C directly onto the gas chromatograph SPB-5, 30m X 0.53 mm ID 1.5 µm film, fused silica capillary column. The column, at 35^{o}C, was held at that temperature

TABLE 2

Summary of test conditions for conducting sediment spiked Jet-A MFC

Organisms	
Organisms per chamber	50 ml as inoculated from stock culture
Experimental design	
Test type	Multispecies
Test vessel	1 l borosilicate glass beakers covered with a 150 x 15 mm petri dish covers
Medium Volume	1 l total
Replicates x concentrations	6 x 4
Concentrations	0 µl, 2 µl, 10 µl, and 25 µl
Addition of toxicant	Day 0
Sampling frequency	Twice weekly
Duration	60 days
Physical and chemical parameters	
Temperature	20 ^{o}C +/- 1 ^{o}C
Light Intensity	Enough to sustain algal growth
Photoperiod	12 hours light/12 hours dark
Medium	900 ml T82MV 8 µg/l $NaHCO_3$
Sediment	50 ml of coadapted plus 50 ml of spiked sediment
Measurements	Dissolved oxygen pH Turbidity Organism counts Gas chromatography Sediment TOC
Parameters Calculated	Photosynthesis (P) Respiration (R) P/R ratio Absorbance Organism densities Total algae Total Protozoa Total Invertebrates

for 2 min., increased to 225°C at 12°C/min and held at that temperature for 5 min. A spectra-Physics 4290 Integrator was used to record the FID signal output of the volatile hydrocarbons that were separated and eluted from the column by molecular weight.

Bacteria

Bacteria were also enumerated each sampling day using an established direct count procedure (Coleman, 1980; Porter and Feig, 1980; Francisco et al. 1973). The epiflourescent stain DAPI (4'6-diamidino-2-phenylindole) was used due to it's superior performance staining small cells (Coleman 1980), it's specificity for active DNA allowing examination of samples with detritus present (Porter and Feig, 1980), and it's superior storage qualities (Porter and Feig, 1980).

Individual replicates within each treatment group were sampled on a rotational basis to conserve volume. One ml was removed from the approximate center of one chamber of each treatment group and sterilly preserved with 40 µl of 0.2 µm filtered glutaraldehyde. These samples were stored in the dark at 4 °C. Slides were prepared by first assembling a 15 ml, 25 mm filtering apparatus. This was done by first washing the filtering apparatus tower and frit in 2N HCL and rinsing in autoclaved deionized distilled water. A droplet of autoclaved and 0.2 µm filtered deionized distilled water was then placed on the frit and a black Poretics brand 25 mm 0.2 µm polycarbonate filter was placed on the bubble of water. The filtration tower was then secured and vacuum was briefly applied to seat the filter.

To determine cleanliness of the apparatus, a blank was prepared each day by adding 3 ml of sterile, autoclaved deionized distilled water via a 10 ml disposable pipette and 300 µls of 50 µg/ml of 0.2 µm filtered DAPI via a 2 ml disposable pipette to the assembled 15 ml filtering apparatus and allowing this to

sit for five minutes. Vacuum was then applied slowly until all of the water was filtered without allowing the filter to be sucked dry. The filter was then removed from the disassembled apparatus using sterile tweezers and placed on a droplet of Resolve low fluorescence immersion oil on a clean kim-wiped slide. Another drop of oil was then added followed by a 25 mm round no. 1 coverslip using the tweezers. This was then tapped with the tweezers to push out any air bubbles. Finally, the slide was labelled and viewed with a Nikon Otiphot-2 epiflourescent microscope at 1250X using Resolve immersion oil under the lens. Fifteen fields were examined under UV for contamination using a 71 x 71 µm Whipple eyepiece grid. If less than one - two cells were seen per field the slide was accepted as a blank. The slide was then wrapped in foil and placed in a plastic labelled slide storage box and stored at 0 ^{o}C.

Sample slides were then prepared for viewing in a similar fashion except they were vortexed well prior to having 200 µl removed and placed in the filtering tower along with 200 µl of 50 µg/ml 0.2 µm filtered DAPI and 1800µl of the autoclaved deionized distilled water. Enumeration and recording of the total bacterial cell counts of 15 fields of the 71 x 71 µm Whipple eyepiece grid using epiflourescent microscopy was then done for each sample slide. The slides were then immediately foil wrapped to exclude light and stored in a labelled plastic slide box at 0 ^{o}C.

Total Organic Carbon

At the conclusion of the microcosm experiment on day 60, all chambers were drained using a siphon hose and allowed to air dry. The flocculent upper sediment layer was removed during siphoning and was discarded. All air-dried microcosm sediments were then stirred with sterile glass rods and shaken for 30 seconds on a vortex mixer. The contents were then blended for two minutes in a small blender. Approximately 10 ml from each beaker was then placed in a 30

ml scintillation vial that had been ashed for 12 hours in a muffle furnace. These were then covered with foil and a teflon cap and labelled with treatment group and replicate and shipped to a lab[5] to be analyzed for total carbon using a CHN analyzer.

Data Analysis

Calculated Parameters

Data from the MFC was recorded on computer entry forms and subsequently entered into a computer. Entries were checked for accuracy and numerical densities of each monitored category were calculated along with net photosynthesis (P), respiration (R), photosynthesis/respiration ratio (P/R), absorbance (A), and total algae in accordance with the SAM protocol (ASTM E 1366-91). Total protozoa and total invertebrates were calculated in a manner similar to total algae. Numerical densities of each algal and protozoan category were calculated as cells or organisms per ml. The "large" organisms *H. azteca*, copepods and ostracods were calculated as organisms per 100 ml.

Net photosynthesis (P) was calculated as follows:

$$P = PMDO2 - AMDO1$$

Where: AMDO1 = first a.m. measurement
PMDO2 = first p.m. measurement

and is simply the net photosynthesis for the daytime period.

Night respiration (R) was calculated as follows:

$$R = PMDO2 - AMDO3$$

Where: AMDO3 = second a.m. measurement

and is simply net dark period respiration.

The P/R ratio was calculated as follows:

---

[5]Katherine Ann Krogslund, Senior Oceanographer, Manager, Marine Chemistry Lab, 224 Old Oceanography Building WB-10, University of Washington, Seattle, WA 98195.

$$P/R = (PMDO2-AMDO1)/(PMDO2-AMDO3)$$

Values greater than/less than one indicate a net oxygen/biomass gain/loss for the 24 hour period.

Absorbance was calculated as:

$$-\log_{10}(\text{percent transmission}/100)$$

and is simply a physical measure of the light absorbance of the media.

Total algae was calculated as:

$$\Sigma Algae_i \ (10^3 \text{ cells/ml})$$

which is simply the sum of all algal cells, expressed as 10^3 cells per ml.

Total protozoa was calculated as:

$$\Sigma Protozoa_i \ (\text{organisms/ml})$$

which is simply the sum of all protozoa, expressed as organisms per ml.

Total invertebrates was similarly calculated as simply organisms per 100 ml:

$$\Sigma Invertebrate_i \ (\textit{organisms/100 ml}).$$

Bacterial cells, in order to be more consistent with the goals of the test, and to avoid the large statistical error involved with calculating cells per unit volume (usually cells per l), were simply reported as the number of cells per 15 fields as generated from the direct epiflourescent counts.

Univariate Statistics

The statistical significance of most of these calculated parameters, along with the physical parameter data for dissolved oxygen and pH, were computed using the Interval of non-significant difference (Conquest and Taub, 1989). ANOVA's were calculated each sampling day for each variable and were used to plot average daily values and IND's over time in order to identify significant differences between the controls and treatments under a null hypothesis of no treatment effect:

$$IND = mean_{oulgroup} \pm t_{d.f.} \sqrt{MSW(1/n_t + 1/n_c}$$

where:

$t_{d.f.}$ = Students t value for the degrees of freedom associated with the mean square (MS) error term of ANOVA.

MSW = Mean Square (MS) within group error term from daily ANOVA.

n_t and n_c = number of treatment/control replicates.

Each of the data sets for the monitored biological variables was first transformed ($\log_{10} + 1$) before calculating the ANOVA and the IND to allow for nonnormal distributions of the data. Values were then transformed back ($10^x - 1$) into original values and these IND's were plotted with original means of the untransformed data.

Bacterial cells and total organic carbon, due to the limitations on the data, were handled differently. Bacterial cells, due to the absence of replication, were analyzed using a Pearson's correlation matrix under a null hypothesis of each comparison being uncorrelated ($\rho=0$). Linear regression was also used regressing each treatment group against time in order to compare slopes for dose-response relationships under the assumption that cell numbers were dependent upon treatment. These were conducted under a null hypothesis of no linear relationships between time and cell numbers.

Total organic carbon, due to the absence of temporal data, was simply tested for normality, homogeneity, and subjected to ANOVA under a null hypothesis of treatment having no effect.

Multivariate Statistics

Three multivariate significance tests were also used. Two of these used the distance measures of cosine of the vector and Euclidean distance between test chambers (Good, 1982; Smith et al, 1990). Statistical significance was determined by analyzing the average within and between group distances using a permutation test (Noreen, 1989). The third test, RIFFLE, utilizes a nonmetric clustering algorithmn (Matthews and Hearne, 1991) and a simple observed -

expected contingency χ^2 goodness-of-fit association analysis to determine the significance of the clustering produced by Riffle.

Due to the suspected dependence of the functional variables on the structural variables, multivariate analysis was performed on functional and structural components separately. In addition to this, the derived variables of P, R, P/R ratio, total algae, total protozoa and total invertebrates were excluded due to their dependence upon other, included variables. The parameters used for multivariate analysis are listed in Table 3.

For the Euclidean distance and cosine of the vector tests, individual replicates were treated as a vector of values with one value for each measured parameter:

$$X = (x_1 \ldots x_n)$$

Euclidean distance between replicates was then computed as:

$$\sqrt{\sum_i (x_i - y_i)^2}$$

Where x and y are values for each measured parameter from each of two compared replicates.

Similarly, the cosine of the vector between replicates was computed as:

$$1 - \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2 \sum y_i^2}}$$

A ratio test of average within group distances over average between group distances (W/B), analogous to ANOVA, was then used to determine significance of the groupings (Smith et al., 1990). For example, for each sampling date, the average within/between group ratio was computed as:

$$\frac{\text{mean within group distance}}{\text{mean between group distance}} = \text{W/B ratio}$$

A large ratio indicates relatively larger distances for within groups as opposed to between groups indicating a poor within treatment clustering effect. A small ratio, on the other hand, indicates a smaller within group distance compared to

between groups, indicating more of a treatment effect. The significance of this grouping was then determined using an approximate randomization test (Noreen, 1989):

$$\text{probability (p)} = (n + 1)/(500 + 1)$$

where n = the number of times a ratio less than the actual within/between ratio is obtained.

This test essentially reassigns replicate labels and recomputes the W/B ratio a large number of times (500). The value obtained is analogous to a statistical probability value "p" and if a larger ratio, on average, is obtained more than 95% of the time, the test is considered significant at the $\alpha = 0.05$ level under a null hypothesis of no treatment effect. These significance levels were then plotted over time.

In the nonmetric clustering analysis Riffle (Matthews and Hearne, 1991), the data were first clustered independently of treatment group. Clusters generated by Riffle may, or may not correspond to treatment group. The null hypothesis for this procedure states that treatment groups and cluster numbers have no association. To evaluate whether clusters assigned by the program corresponded to treatment group, a simple Pearson's χ^2 observed - expected 4 x 4 contingency goodness-of-fit test (Fienberg, 1985) was conducted:

$$\chi^2 = \sum_{ij} \frac{(N_{ij} - n_{ij})^2}{n_{ij}}$$

where N_{ij} is the actual cell count and n_{ij} is the expected cell frequency. The significance (probability) for this value of χ^2 was computed using a standard procedure (Press et al., 1990). Significance levels from this association analysis were then plotted over time.

TABLE 3

Parameters used in the multivariate statistical tests. Derived variables were not used since they are derived from and therefore not independent of other, independent variables.

	Riffle	Euclidean distance & Cosine of the vector
Functional:	pH	pH
	DO1	DO1
	DO2	DO2
	DO3	DO3
	Absorbance	Absorbance
Structural:	*Selenastrum sp.*	*Selenastrum sp.*
	Chlorella sp.	*Chlorella sp.*
	Scenedesmus sp.	*Scenedesmus sp.*
	Ankistrodesmus sp.	*Ankistrodesmus sp.*
	Other Green algae	Other Green Algae
	Filamentous Green	
	Nitzchia sp.	*Nitzchia sp.*
	Other Diatoms	Other Diatoms
	Lyngbya sp.	*Lingbya sp.*
	Other Blue-Greens	Other Blue-Greens
	Amoeba	Amoeba
	Ciliates	Ciliates
	Flagellates	Flagellates
	Paramecium bursaria	*Paramecium bursaria*
	Rotifers	Rotifers
	Hyalella azteca	*Hyalella azteca*
	Copepods	Copepods
	Ostracod 1	Ostracod 1
	Ostracod 2	Ostracod 2
	Insect larvae	
	Bacteria	

RESULTS

Gas Chromatography

Unfortunately, many of the samples taken for gas chromatographic analysis were lost due to biodegradation of the samples while being stored in the test tubes. Satisfactory results, analyzed within several days of sampling, were obtained for all groups on days zero, four, 17, 18, and 21. For day 25, satisfactory results were obtained for the 10 and 25 µl groups only. All satisfactory analytical results were for actual sampling days except those obtained for day 17.

From these results, a pulsed release of Jet-A from the encapsulated sediment was obtained (Fig. 1). The time required for release from the sediment corresponded to dose (Fig. 2). The Jet-A remained in the test systems for a substantial portion of the duration of the test (Fig. 3). However, no clear exposure duration could be determined.

Acute Tests

Graphical cumulative percent mortality results for the acute amphipod bioassays are shown in Fig. 4. Graphically obtained $LC_{50's}$ for the range-finding and definitive tests were approximately 512 and 263 µl of Jet-A. The results of the range-finding test were not subjected to tests for normality, homogeneity, and definitive statistical analysis due to the lack of variance in the 1000 µl treatment group, and the use of fewer than 4 replicates per treatment group precluding the use of nonparametric tests. Data from the definitive test were found to be normal (Shapiro-Wilk's test: square root transformation at the α =0.01 level) but were also unable to pass a Bartlett's homogeneity of variance test due to the lack of variance for the 25 µl treatment group. The data were

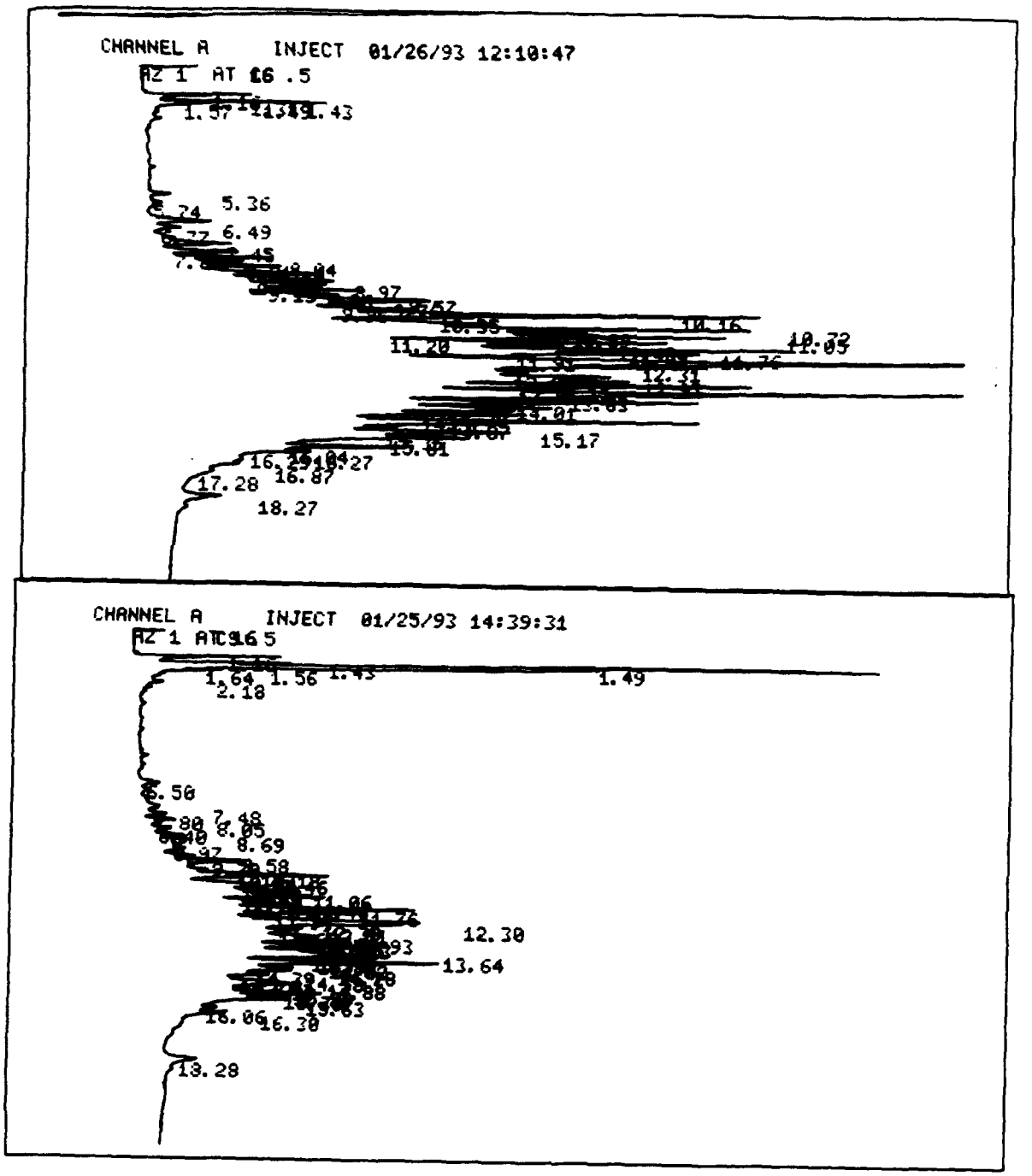


Fig. 1 Gas chromatography results from the 25 µl treatment group on days 17 (1/25/93) and day 18 (1/26/93). Day 17 was taken prior to sampling on day 18. The pulse in concentrations of the Jet-A components is readily apparent.

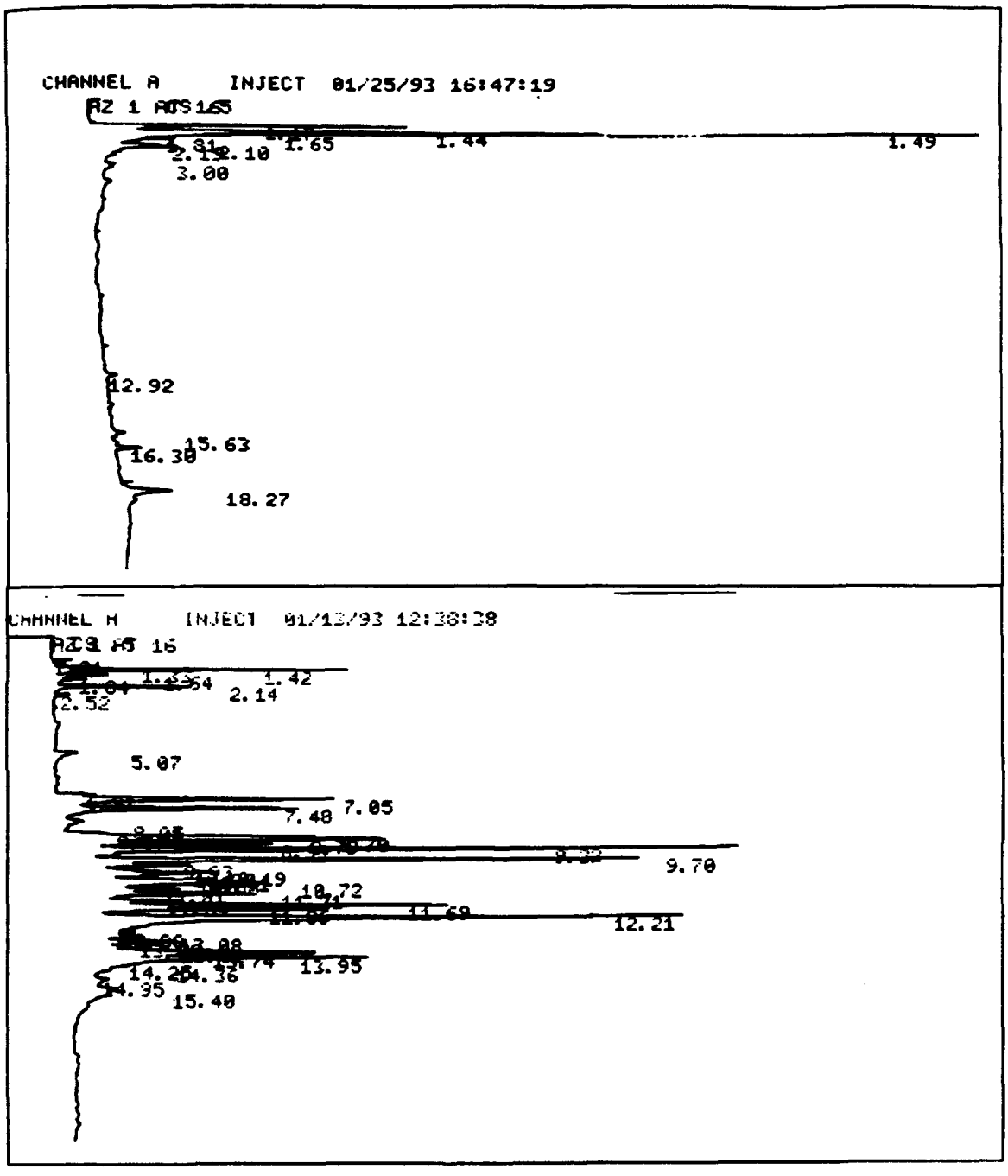


Fig. 2. Gas Chromatography results from days 0 (upper) and 4 (lower)from the 25 µl treatment group. Similar results for other treatment groups on day 4 indicate no release of Jet-A from the sediment had occurred.

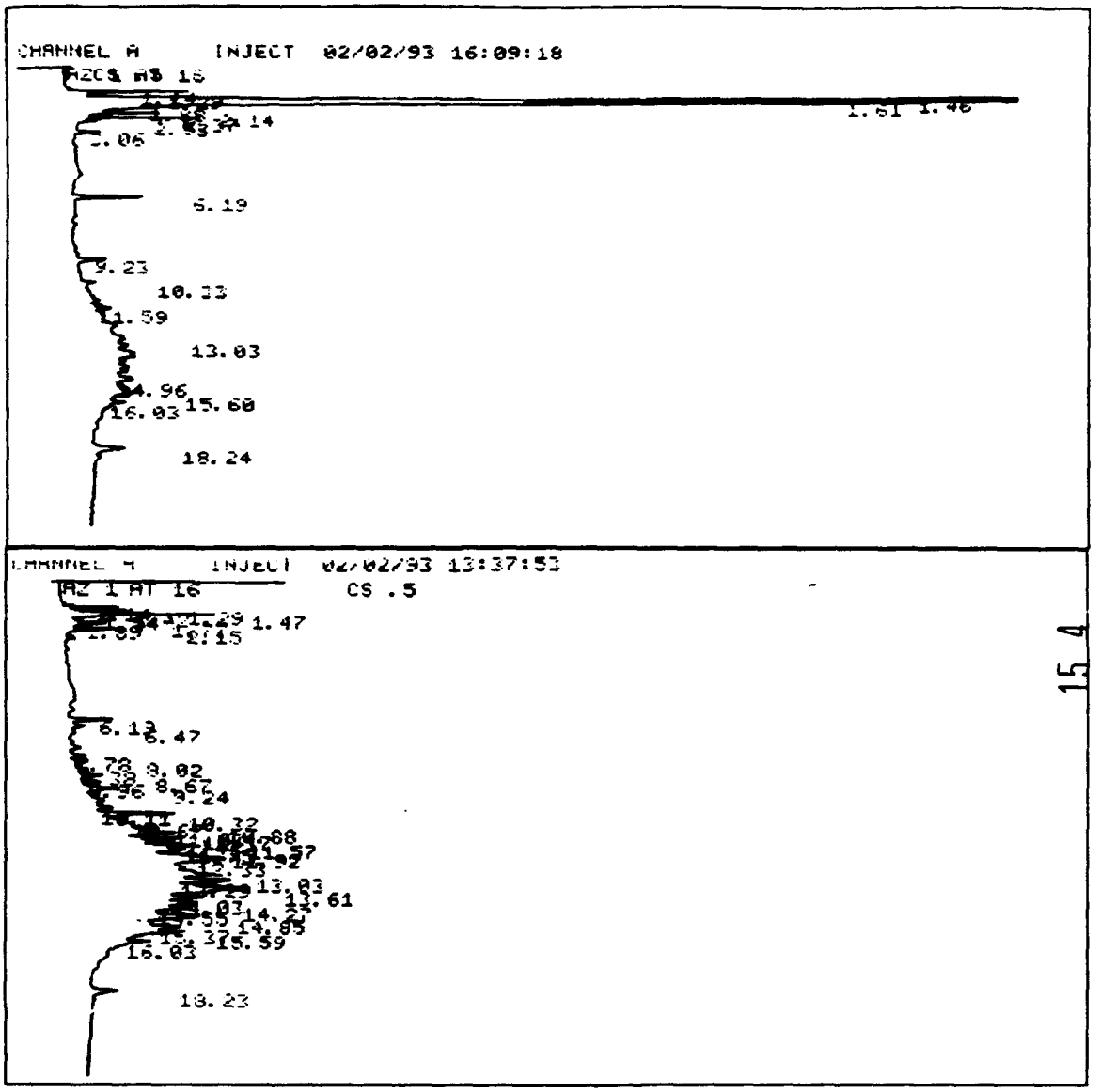


Fig. 3. Gas chromatography results of the 10 µl (upper) and the 25 µl (lower) treatment groups from day 25. Portions of the original spiked Jet-A remain in the test system.

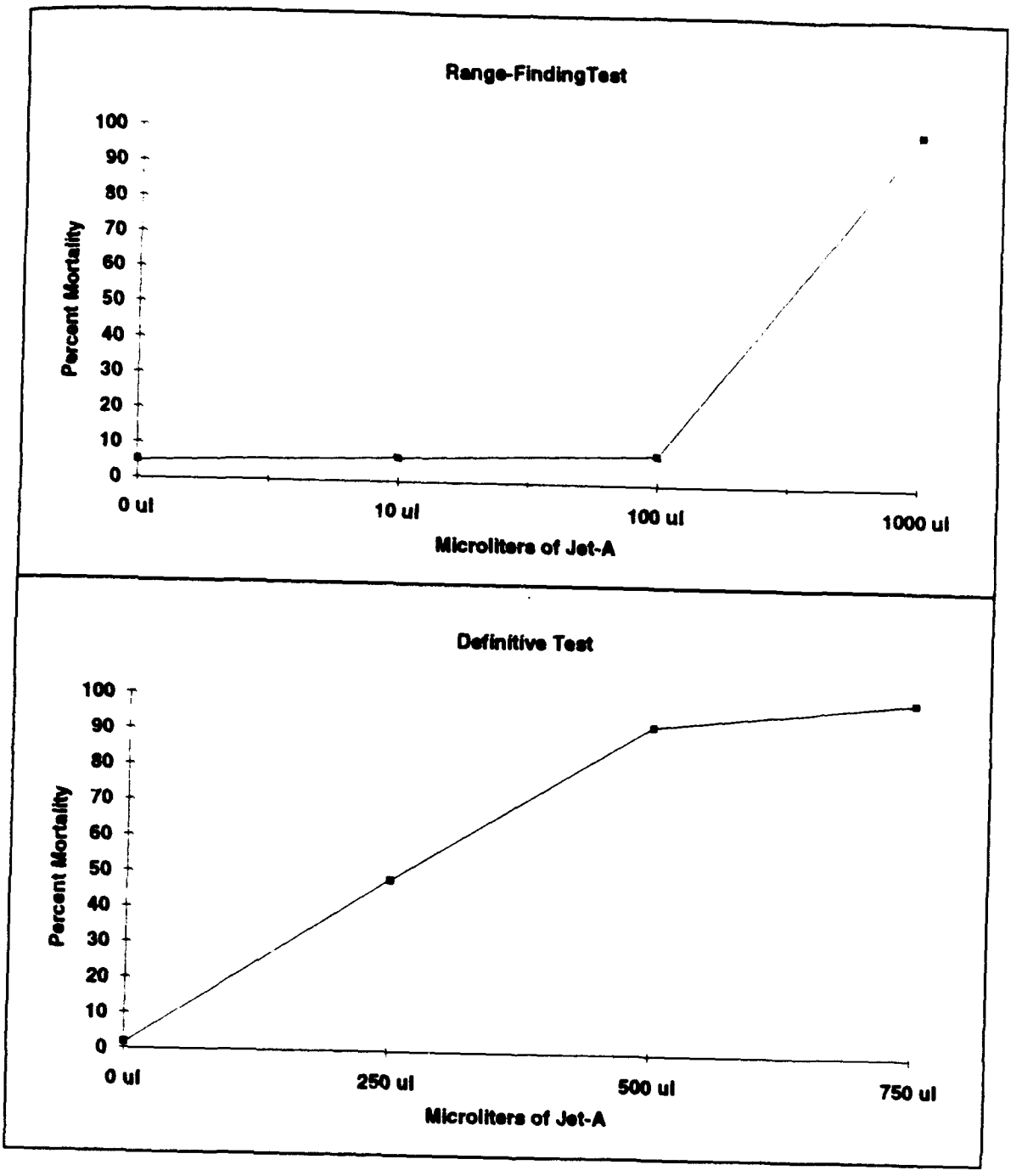


Fig. 4. Hyalella azteca acute percent mortality. The range-finding test concentrations were 0 µls, 10 µls, 100 µls, and 1000 µls. For the definitive test, concentrations were 0 uls, 250 uls, 500 uls, and 750 uls.

subsequently analyzed using ANOVA and found to be significantly different (Ho: all groups equal, square root transformation at $\alpha = 0.05$). Due to the normal distribution of the data, a probit analysis yielded a similar LC_{50} of 259 µls. Water quality characteristics for each test were well within the required parameters for acceptability as specified in the protocol.

Functional and Structural Parameters

Univariate results of the functional parameters indicate that an initial period of perturbation occurred followed by an apparent stable state as defined by the IND. Initial functional dissolved oxygen parameters indicated that an initial period of depression in dissolved oxygen concentrations occurred in all dosed treatment groups on days four through 21 relative to the 0 µl group with all groups generally increasing in concentration over time (Figs. 5-7,11). These were generally statistically significant based on the IND. From days four through seven, a general initial depression was also seen in the 0 µl group (Figs. 5-11). This phenomena was assumed to be due to, in addition to both a stress response to the toxic effect of the Jet-A, and the added heterotrophic substrate provided by the complex mixtures of Jet-A in the dosed groups, an increase in respiration in all treatment groups caused by transfer perturbation increasing the availability of sediment-born heterotrophic substrates. This appeared to occur mainly in the light period and may have been due to the high availability of dissolved oxygen at this time (Figs. 6, 8, 10). On days 14 - 18, a statistically significant increase in the P/R ratio was observed for the 0 µl group, apparently due to a decrease in nighttime respiration during this period (Figs. 7 and 9). Thereafter, a steady general increase in both net photosynthesis and night respiration (Figs. 8 and 9) was observed resulting in an apparent balance in the P/R ratio (Fig. 10).

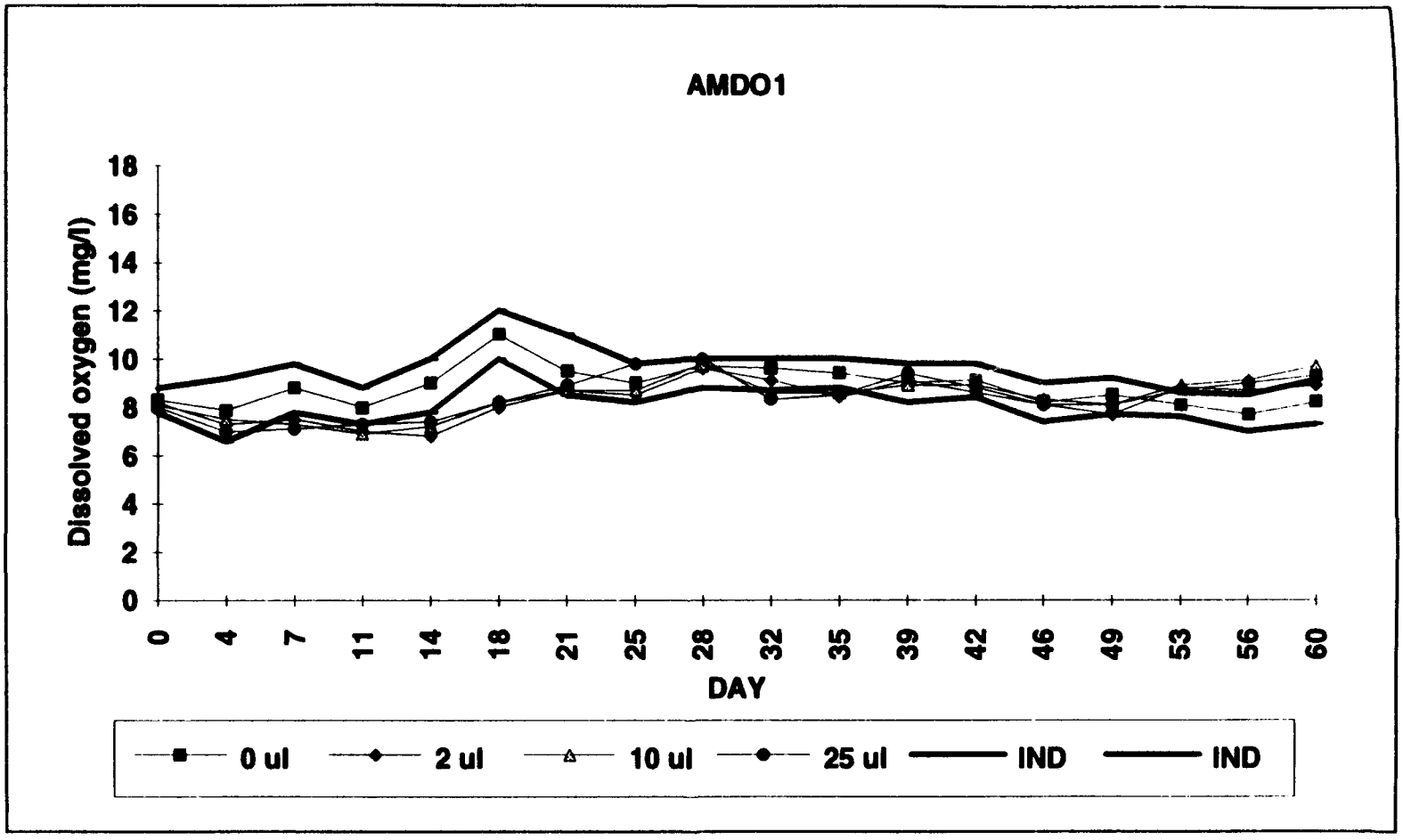


Fig. 5. Changes in AMDO1 over time. First A.M. dissolved oxygen concentrations (mg/l) with the Interval of Nonsignificant Difference (IND).

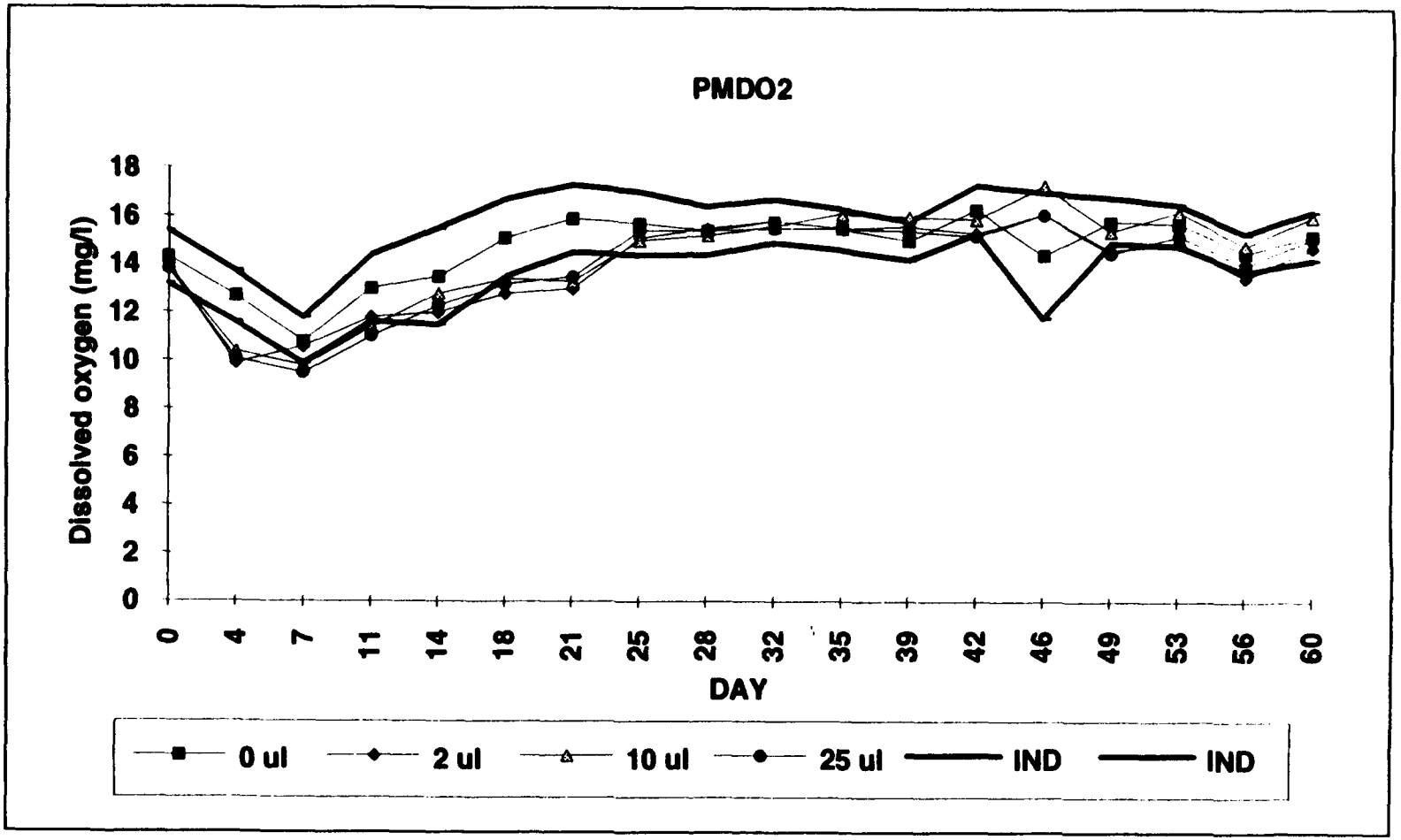


Fig. 6. Changes in PMDO2 over time. First P.M. dissolved oxygen concentrations (mg/l) with the Interval of Nonsignificant Difference (IND).

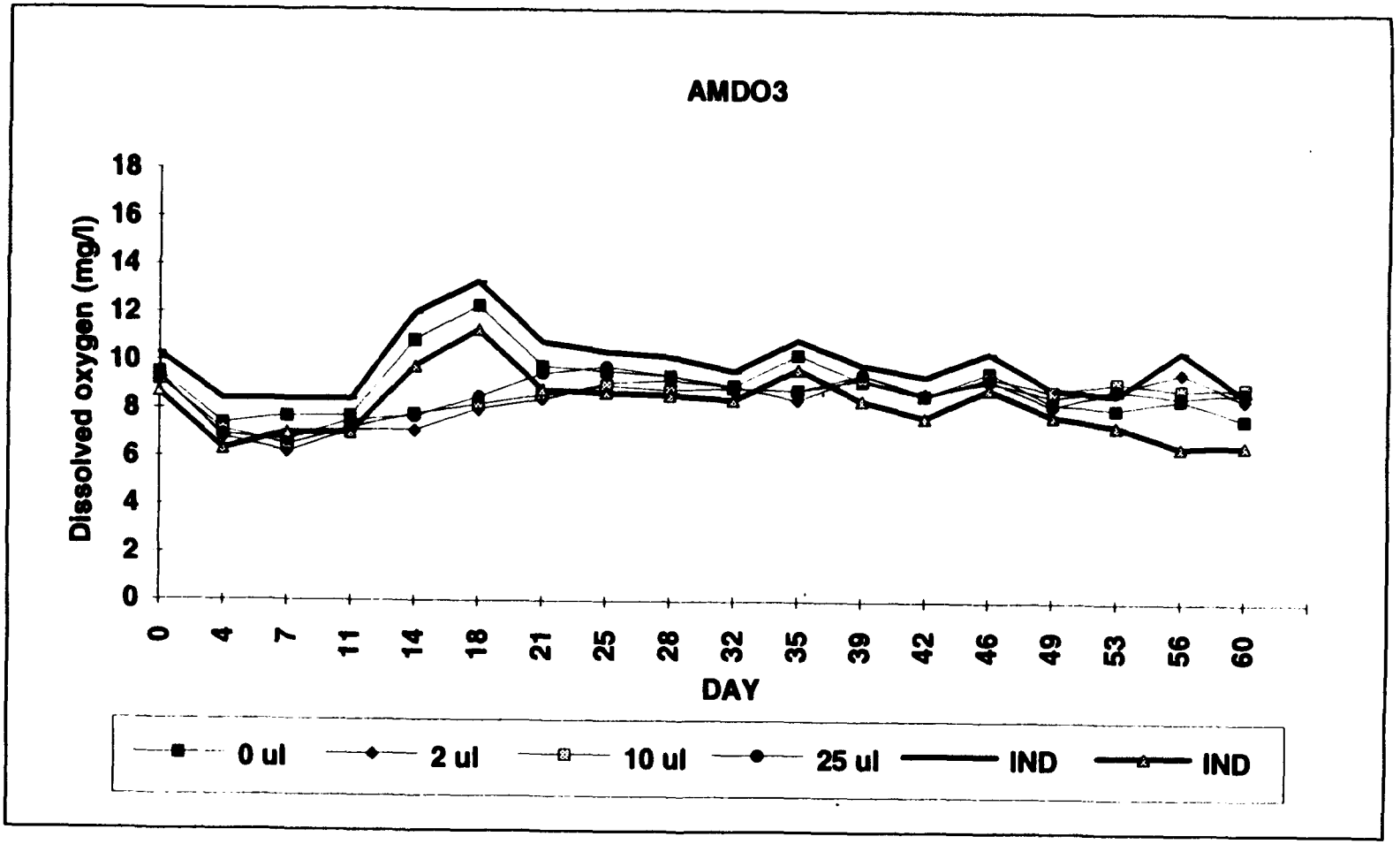


Fig. 7. Changes in AMDO3 over time. Second A.M. dissolved oxygen concentrations (mg/l) plotted with the Interval of Nonsignificant Difference (IND).

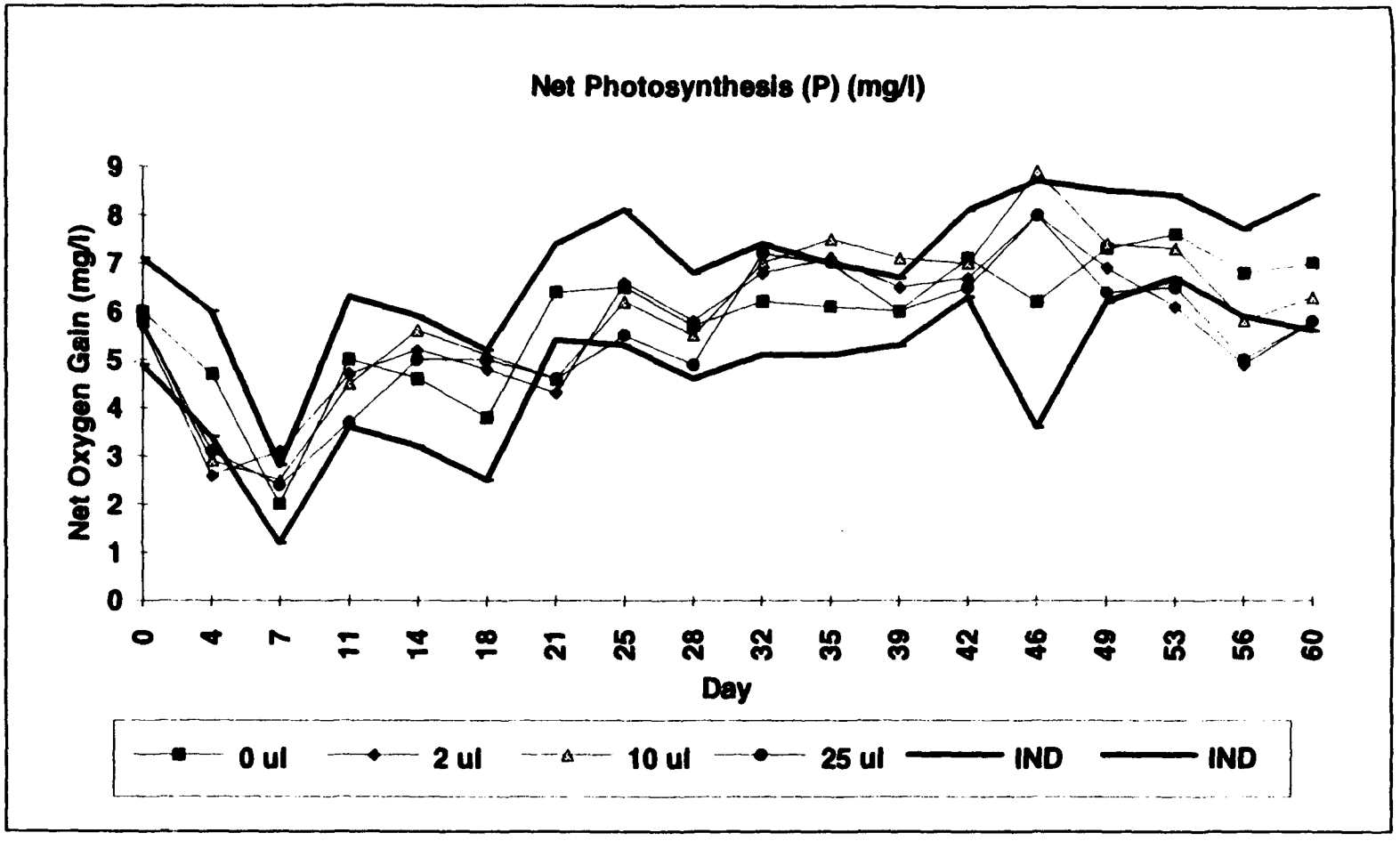


Fig. 8. Changes in photosynthesis (P) over time. The change in net photosynthesis, expressed as net changes in daytime oxygen concentrations, is plotted with the Interval of Nonsignificant Difference (IND). Values are net oxygen gains in mg/l during the daylight period.

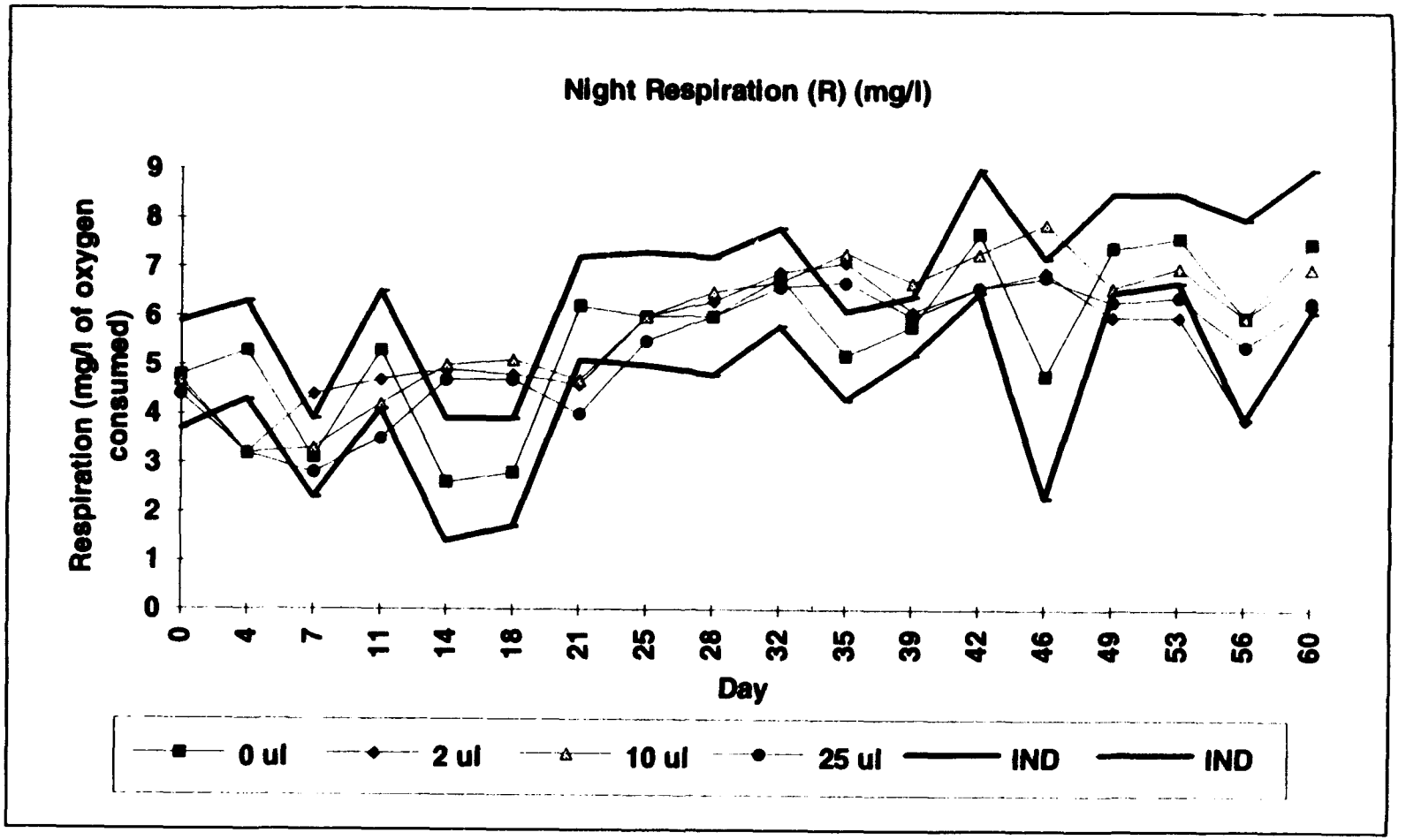


Fig. 9. Changes in respiration (R) over time. The change in net respiration, expressed as net changes in nighttime oxygen concentrations, is plotted with the Interval of Nonsignificant Difference (IND). Values are net oxygen consumption in mg/l during the dark period.

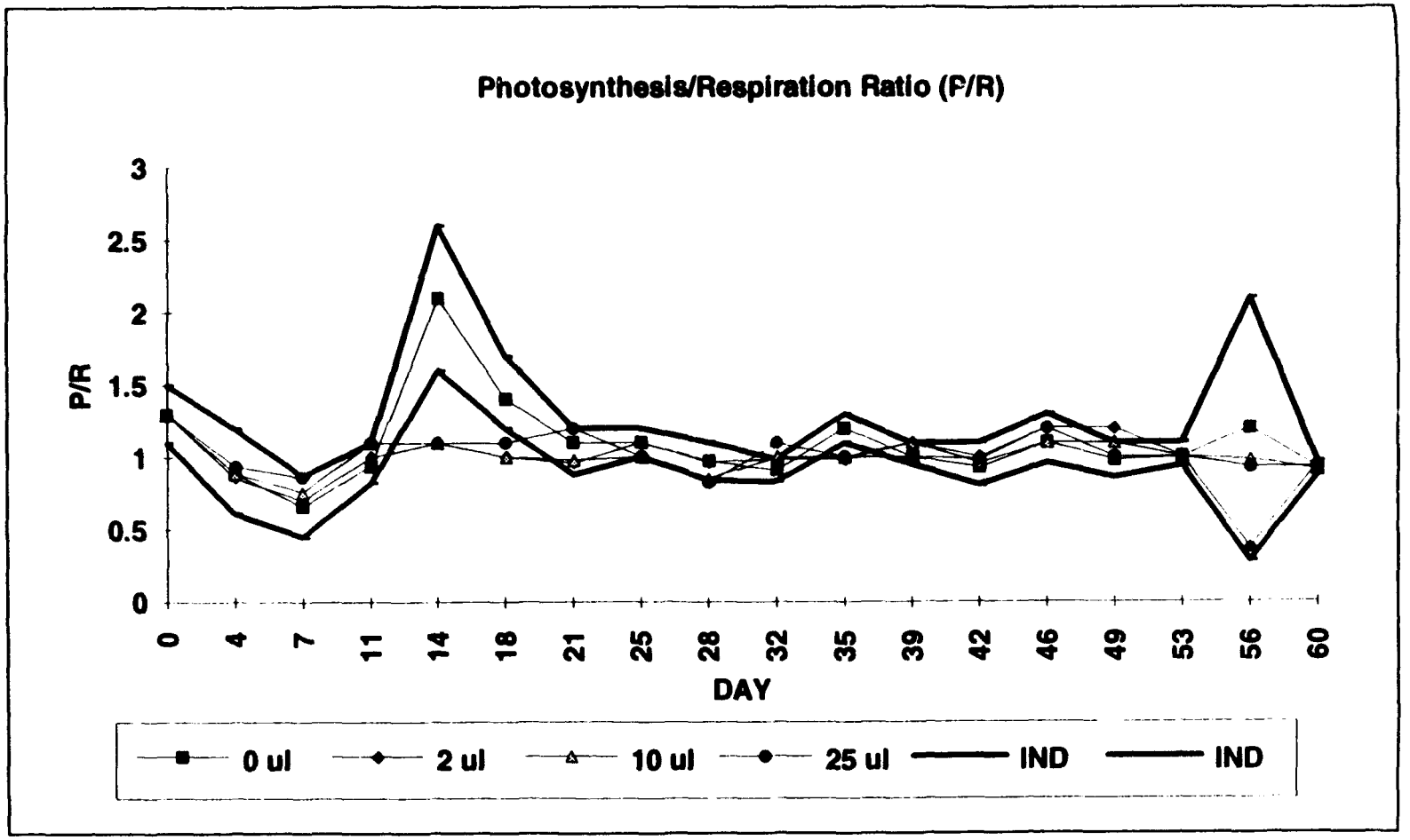


Fig. 10. Changes in the photosynthesis/respiration (P/R) ratio over time. Values greater/less than one indicate a net oxygen or biomass gain/loss over a 24 hour period. Values are plotted with the Interval of Nonsignificant Difference (IND).

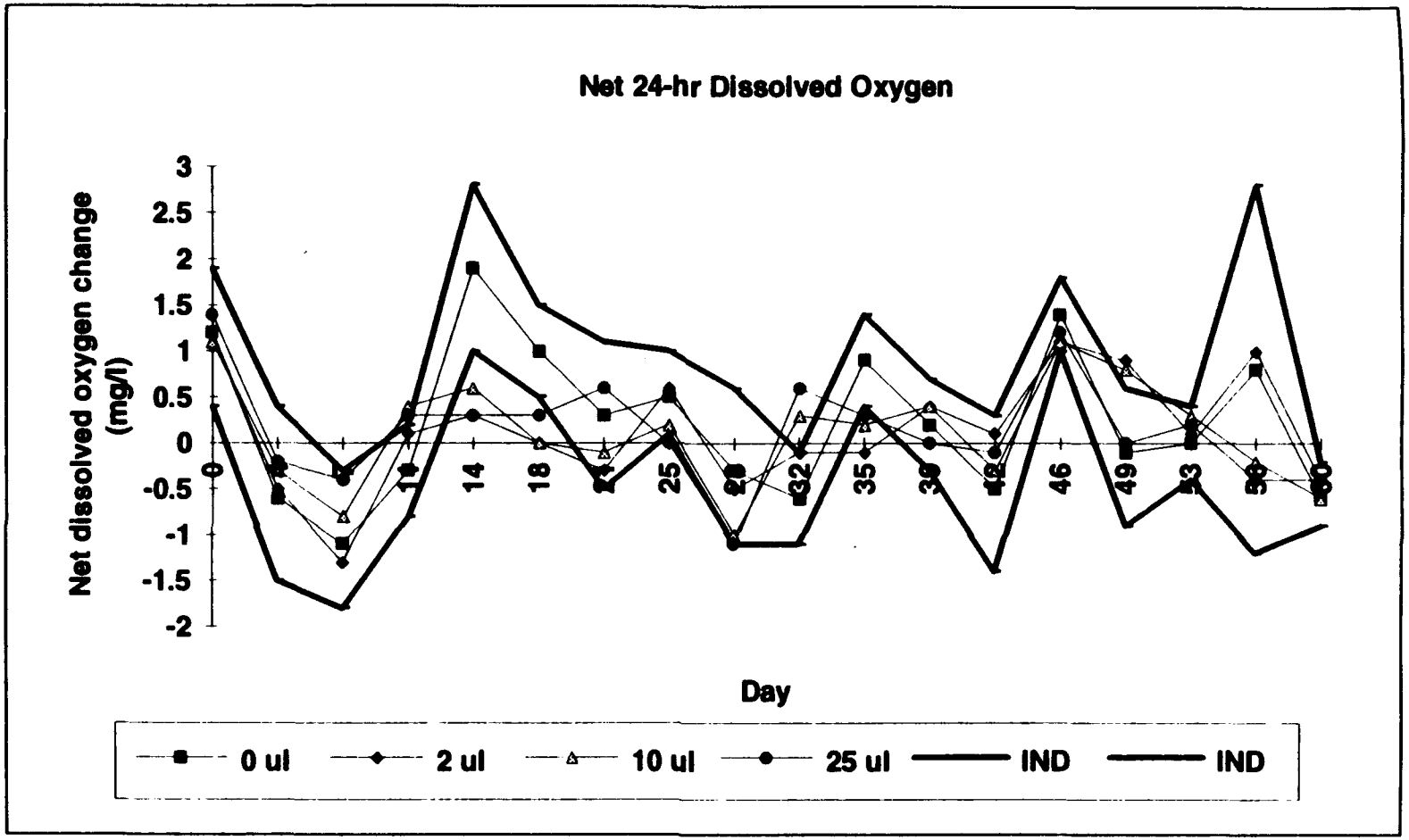


Fig. 11. Net change in 24-hour dissolved oxygen over time. Values represent the net change in AMDO over the 24 hour period preceding each sampling day. Values are plotted with the Interval of Nonsignificant Difference (IND).

The values for the functional parameter pH also agree with this hypothesis. An initial decrease is seen in all treatment groups on days four through seven, due to the predominance of respiration over photosynthesis resulting in a net gain in the hydrogen ion content of the media (Fig. 12). This is statistically significant for all groups on day seven only during this period, presumably due to the nature of the pH variable as essentially log transformed data. Thereafter, a steady increase in pH is seen in all treatment groups, indicating a predominance of photosynthesis over heterotrophic respiration until day 46, when significantly higher pH values were seen for all dosed groups presumably due to the smaller populations of protozoa and invertebrates present.

Absorbance also indicates an increase in photosynthesis on days seven through 28 in the 0 µl treatment group (Fig. 13) and is in agreement with the trend observed with PMDO2 (Fig. 6). This was apparently due to the relatively larger growth rate of *Scenedesmus sp.* and *Anabaena sp.* in the 0 µl group during this period (Figs. 15, 17 and 18). Some statistically significant differences from the 0 µl group were seen during this period based on the IND.

Initially, univariate results of structural parameters indicate that an initial general imbalance in population sizes existed at the beginning of the treatment period on day zero. Thereafter, treatment effects were generally detectable through the entire test, with no apparent stability of the control group or recovery of the system from perturbation.

Total algae (Fig. 14), after an initial imbalance in population sizes on day zero, reveals a steady increase in total algal cells per ml for all treatment groups. Due to within treatment variances, there were essentially never any statistically significant differences between the 0 µl group and all other groups. Of all algal categories monitored, the blue-green categories clearly dominated all treatment groups late in the experiment (Fig. 15). The category Other Blue-Green algae

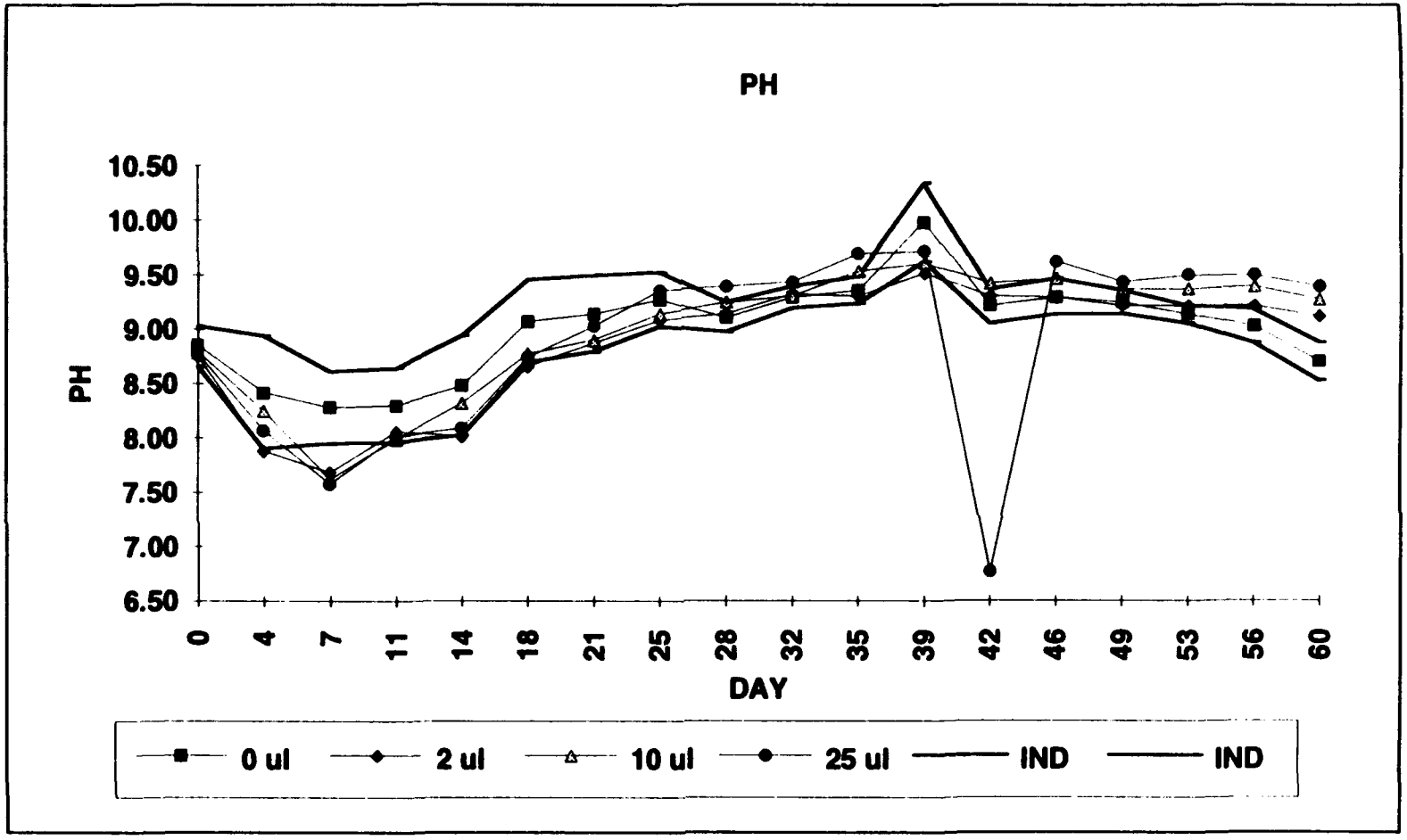


Fig. 12. Change in pH over time. Values are plotted with the Interval of Nonsignificant Difference (IND).

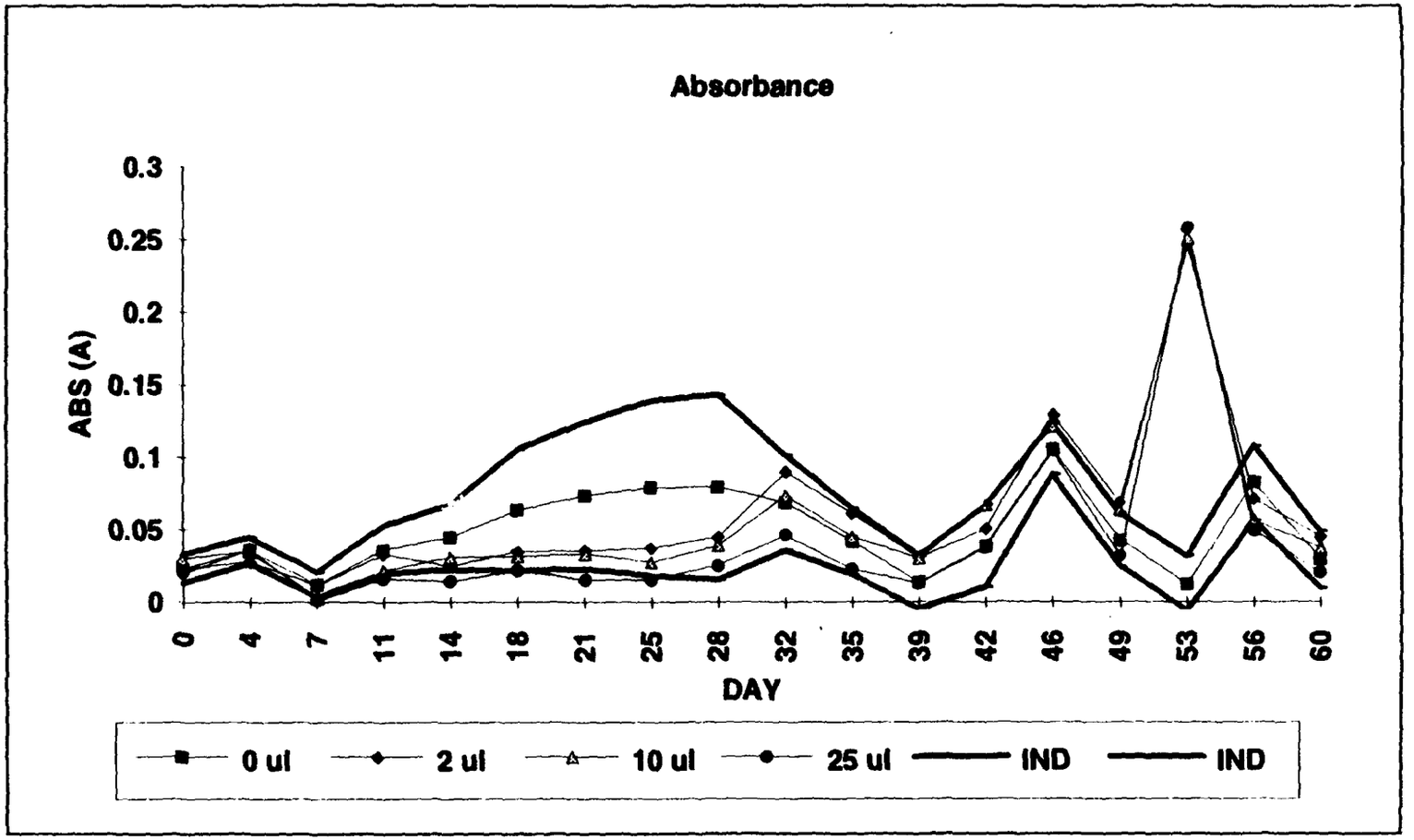


Fig. 13. Changes in absorbance (A) over time with the Interval of Nonsignificant Difference (IND).

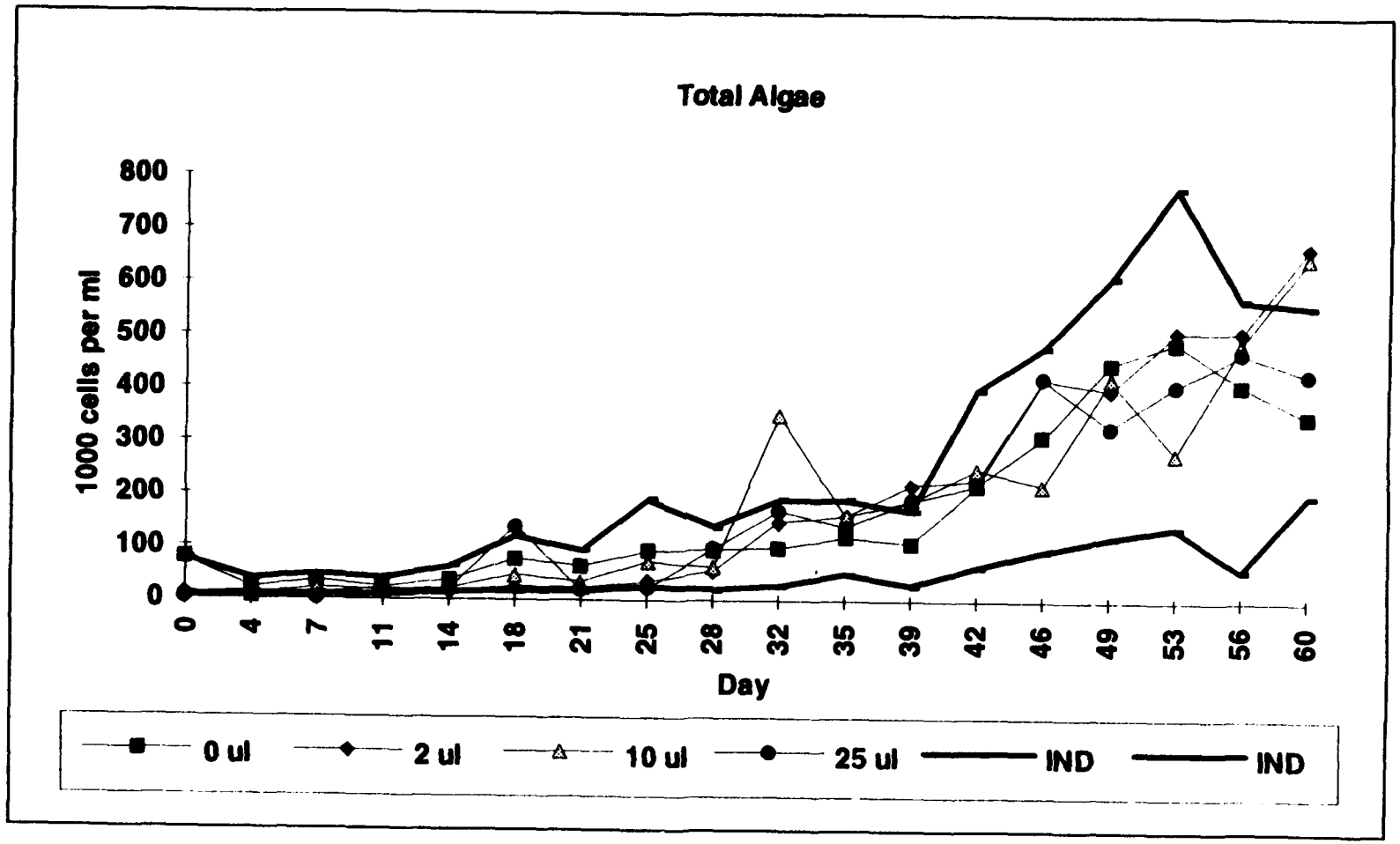


Fig. 14. Changes in total algal cells per ml over time. Values correspond to 10^3 cells per ml and are plotted with the Interval of Nonsignificant Difference (IND).

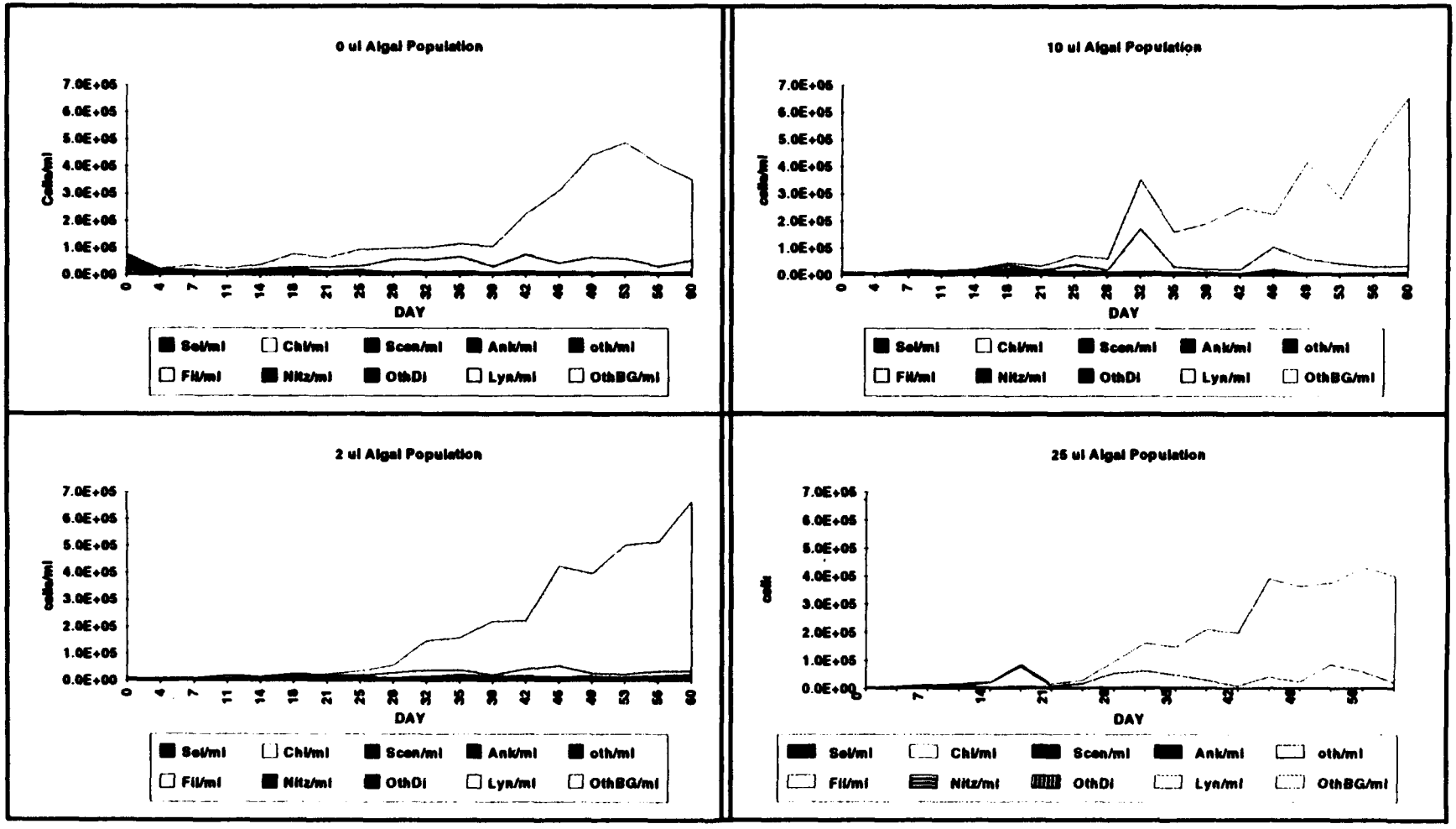


Fig. 15. Changes in algal populations over time. Values for each population are treatment means and are useful for illustrating patterns in population dynamics.

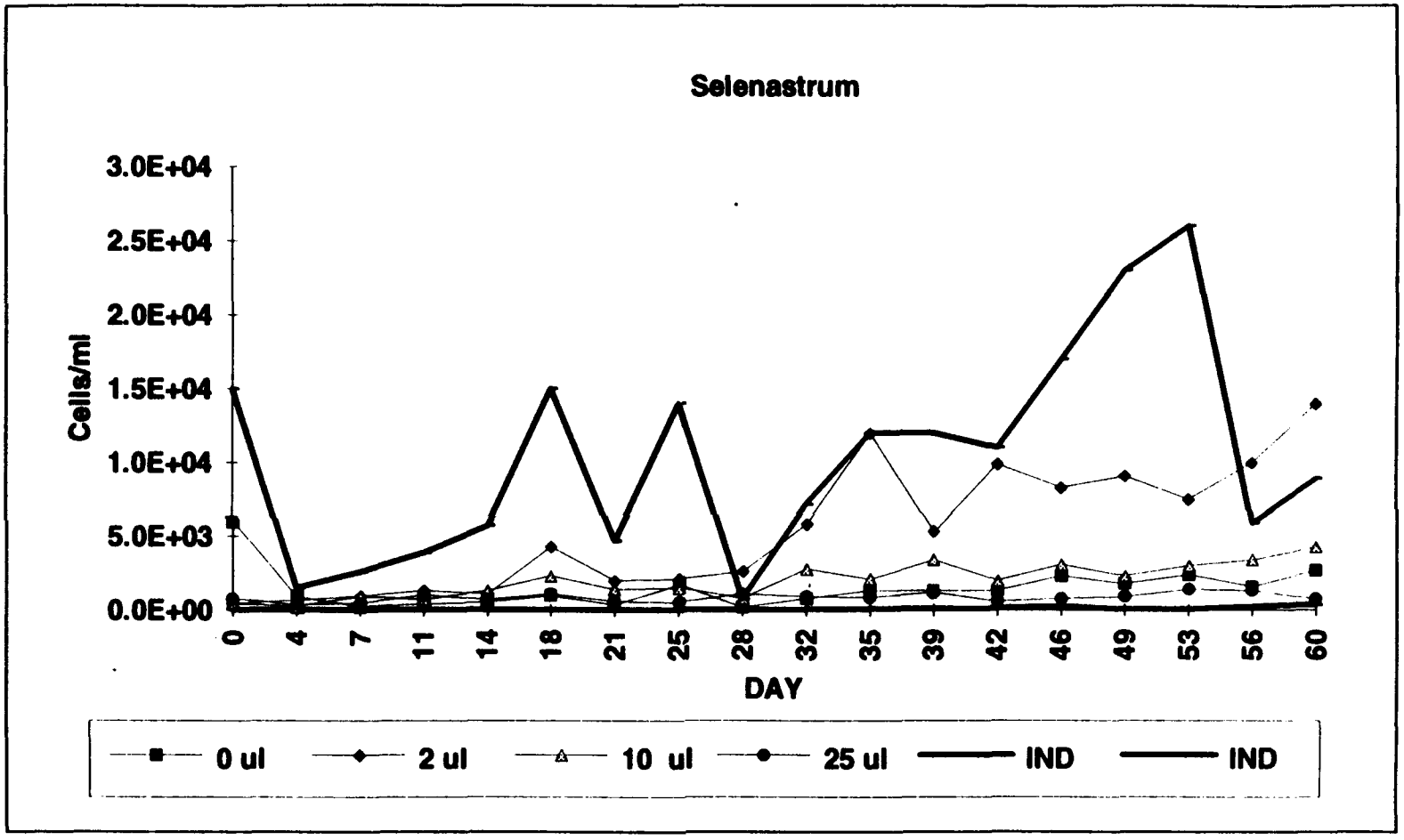


Fig. 16. Changes in densities of *Selenastrum* over time. Values are treatment group means and are plotted with the Interval of Nonsignificant Difference (IND).

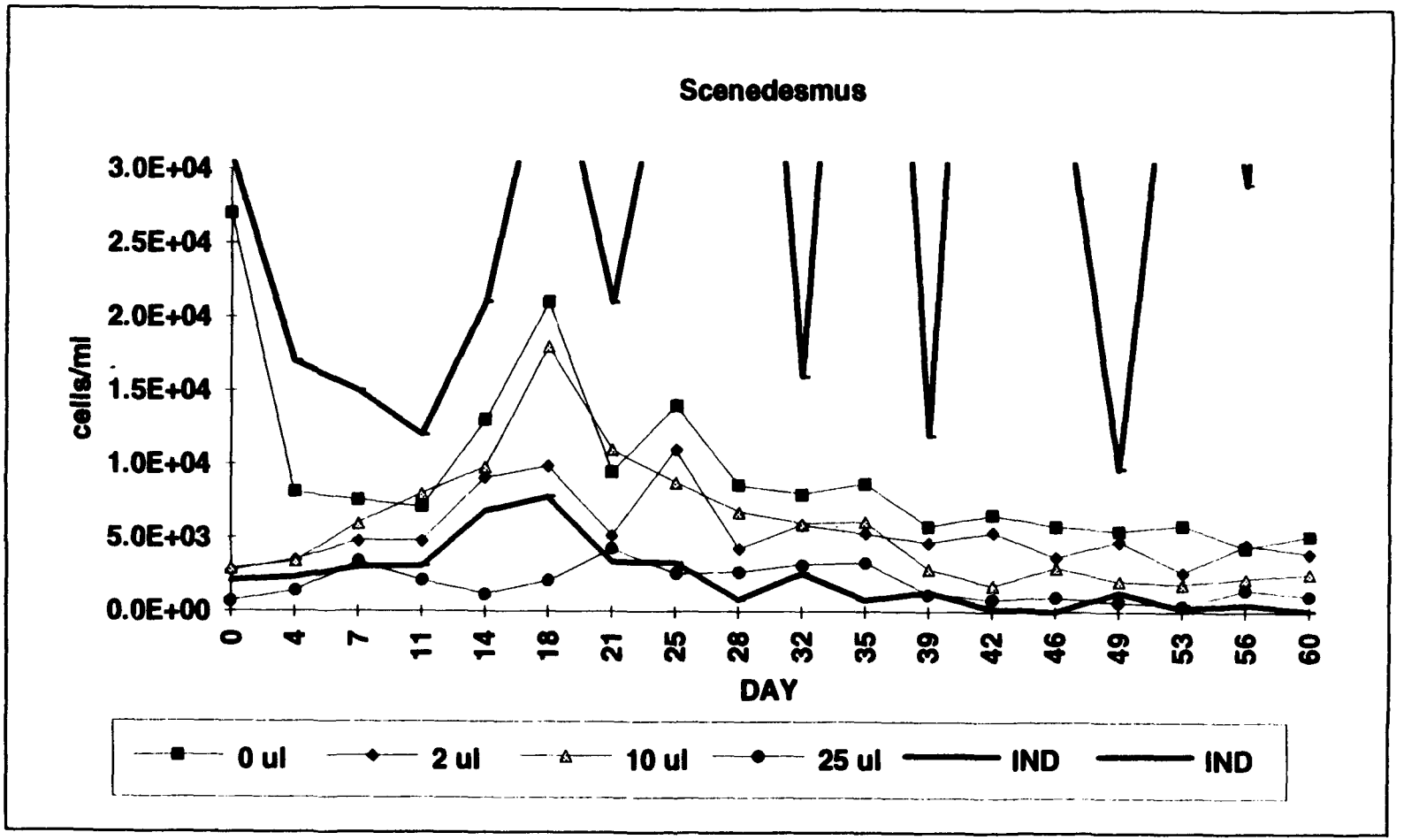


Fig. 17. Changes in densities of *Scenedesmus* over time. Values are treatment group means and are plotted with the Interval of Nonsignificant Difference (IND).

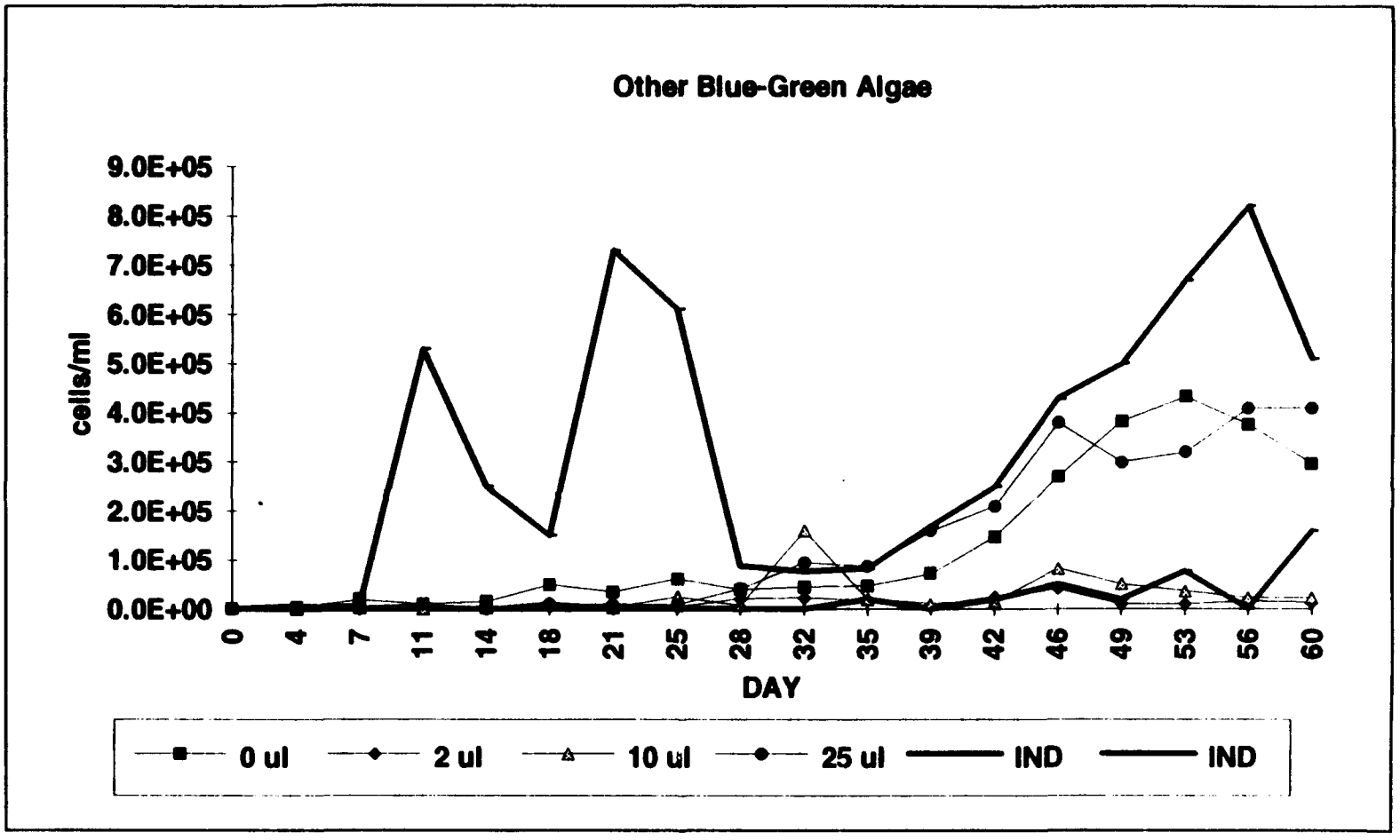


Fig. 18. Changes in densities of Other Blue-Green algae, primarily *Anabaena*, over time. Values are treatment group means and are plotted with the Interval of Nonsignificant Difference (IND).

was essentially comprised entirely of *Anabaena*. No apparent recovery or stability of the 0 µl treatment group was apparent.

Of the individual algal categories monitored, *Selenastrum sp.*, *Scenedesmus*, and Other Blue greens were the most noteworthy. Although not statistically significant, *Selenastrum* appeared to go through a classic algal dose response with the 2 µl group as the most stimulated in growth(Fig. 16). *Scenedesmus* peaked in growth on day 18 for the 0 µl group with the 25 µl group densities significantly lower during much of the experiment (Fig. 17). Other Blue-Green algae, again comprised overwhelmingly of *Anabaena*, were apparently more dense in the 0 and 25 µl treatment groups during the latter stages of the test although this was not statistically significant based on the IND due to a large within group variance (Fig. 18).

Although not statistically significant, similar apparent differences in initial population sizes of total protozoa numbers were observed and instability of the 0 µl group is apparent throughout much of the test (Fig. 19). Although not generally significant based on the IND due to large within group variance, total numbers for all of the dosed treatment groups were consistently less than the 0 µl group until the last few days, when total numbers fell in this group (Figs. 19-20). One notable exception to these observations is *P. bursaria*, which significantly increased dramatically in numbers in the 25 µl group (Figs. 20-21). All other groups were similar in character to Total Protozoa, with a general instability of the 0 µl treatment group and fewer total numbers in the treated groups.

Population dynamics of the invertebrates were similar in pattern in all treatment groups with an initial small population, apparently similarly uneven in size as the algal and protozoan categories (Fig. 22-23). Total numbers appear depressed in both the 10 and 25 µl treatment groups throughout the test with no

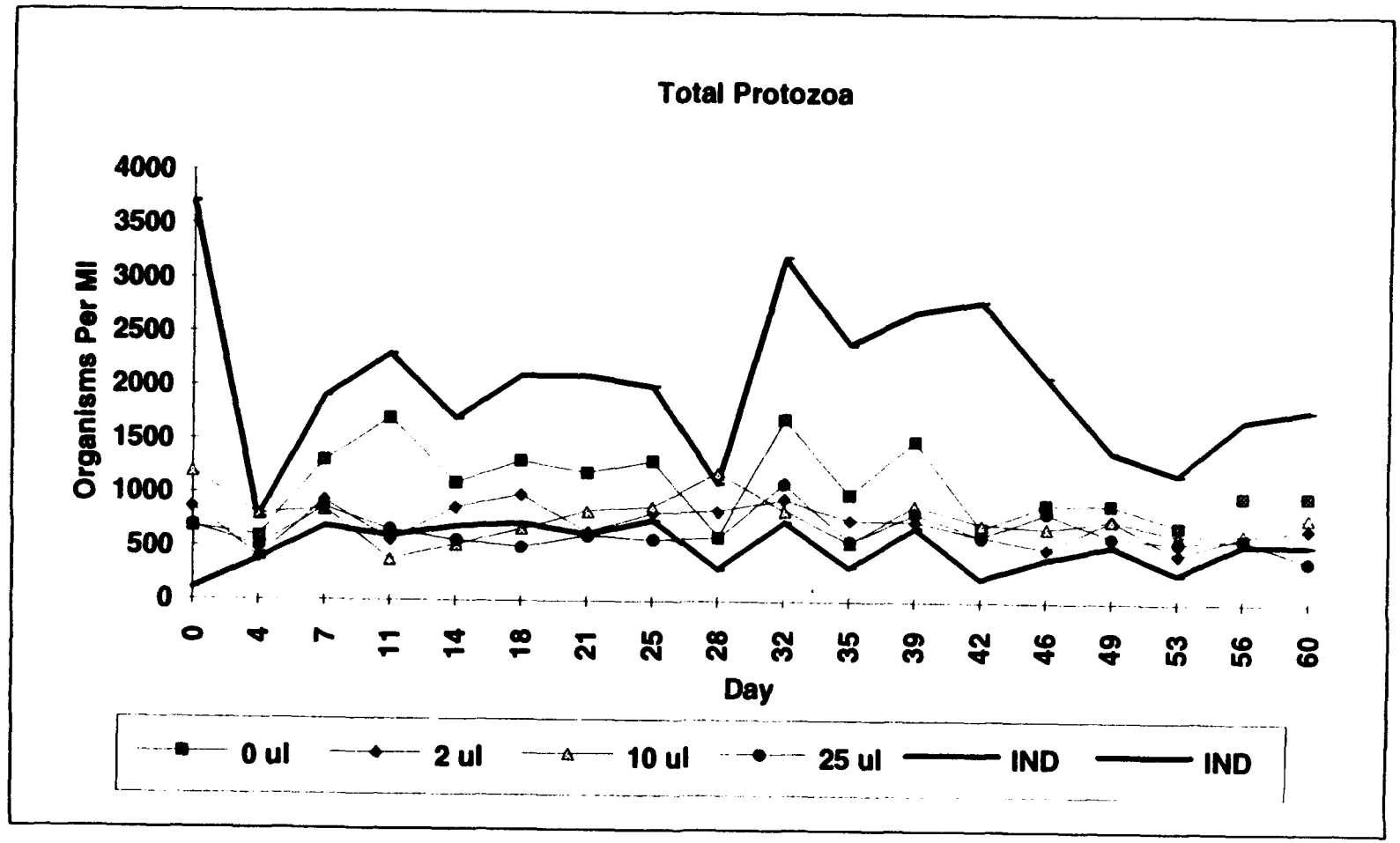


Fig. 19. Changes in total protozoa, as organisms per ml, over time. Values are treatment group means and are plotted with the Interval of Nonsignificant Difference (IND).

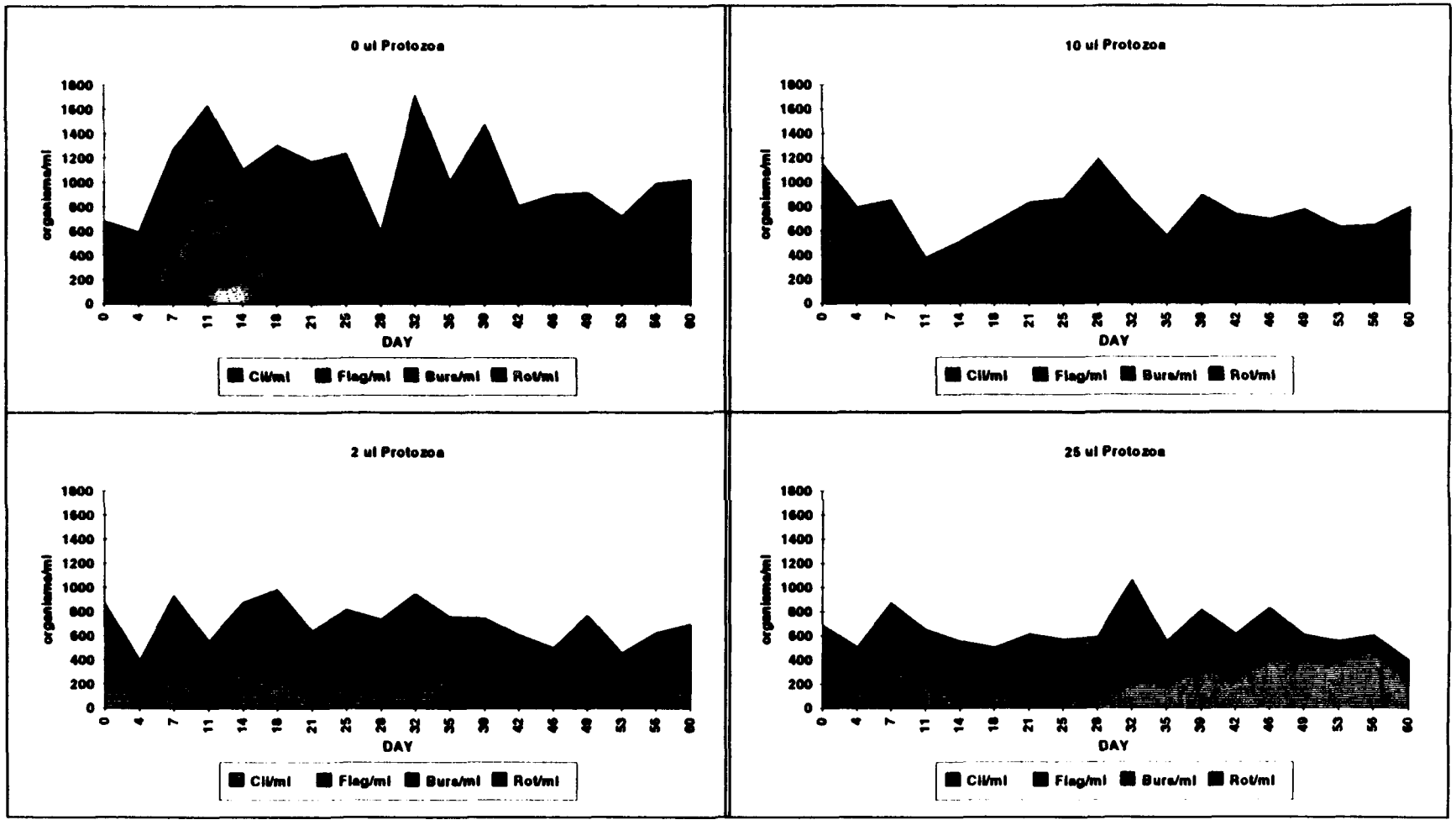


Fig. 20. Protozoa population means for all treatment groups.

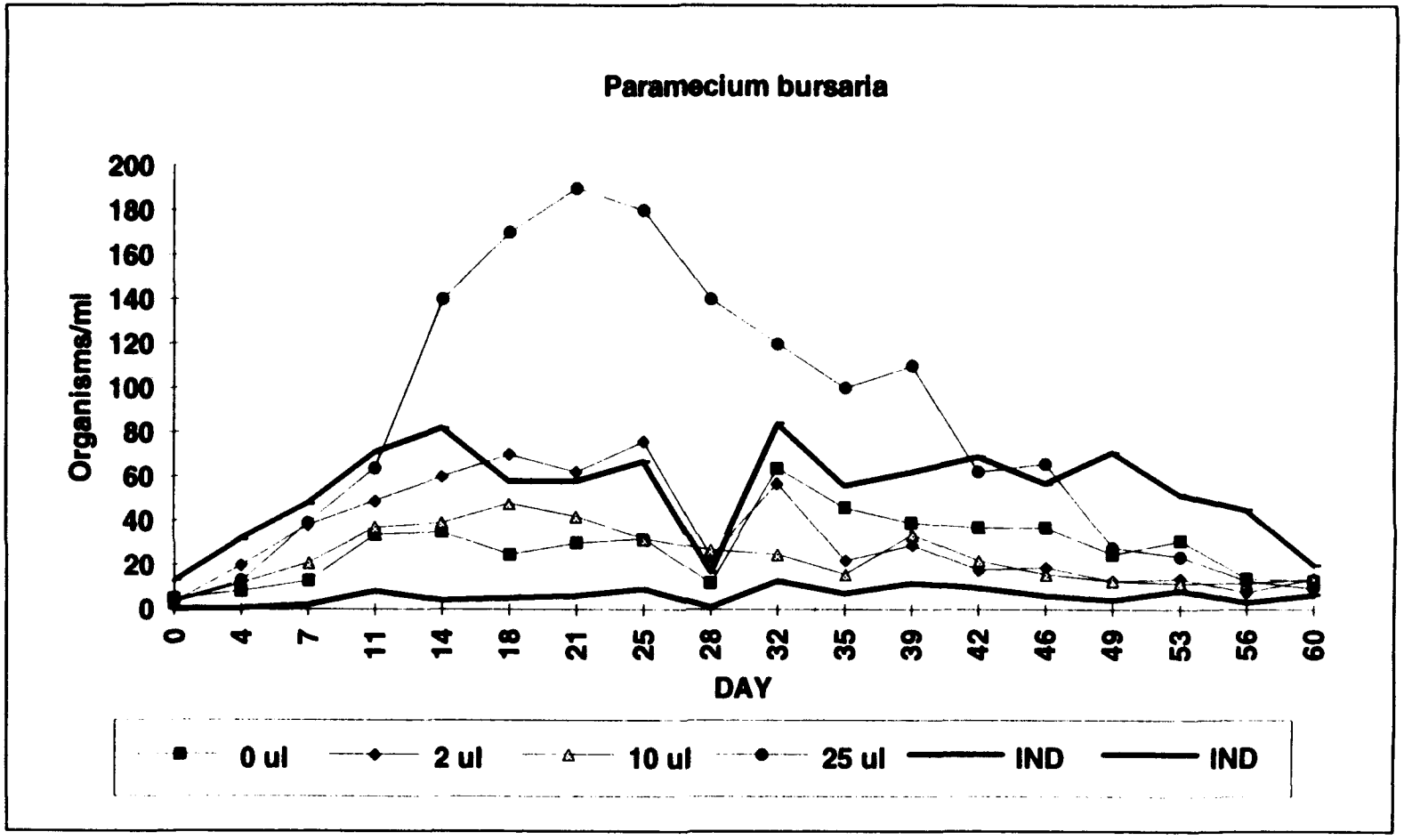


Fig. 21. Changes in densities of *P. bursaria* over time. Values are treatment group means and are plotted with the Interval of Nonsignificant Difference (IND).

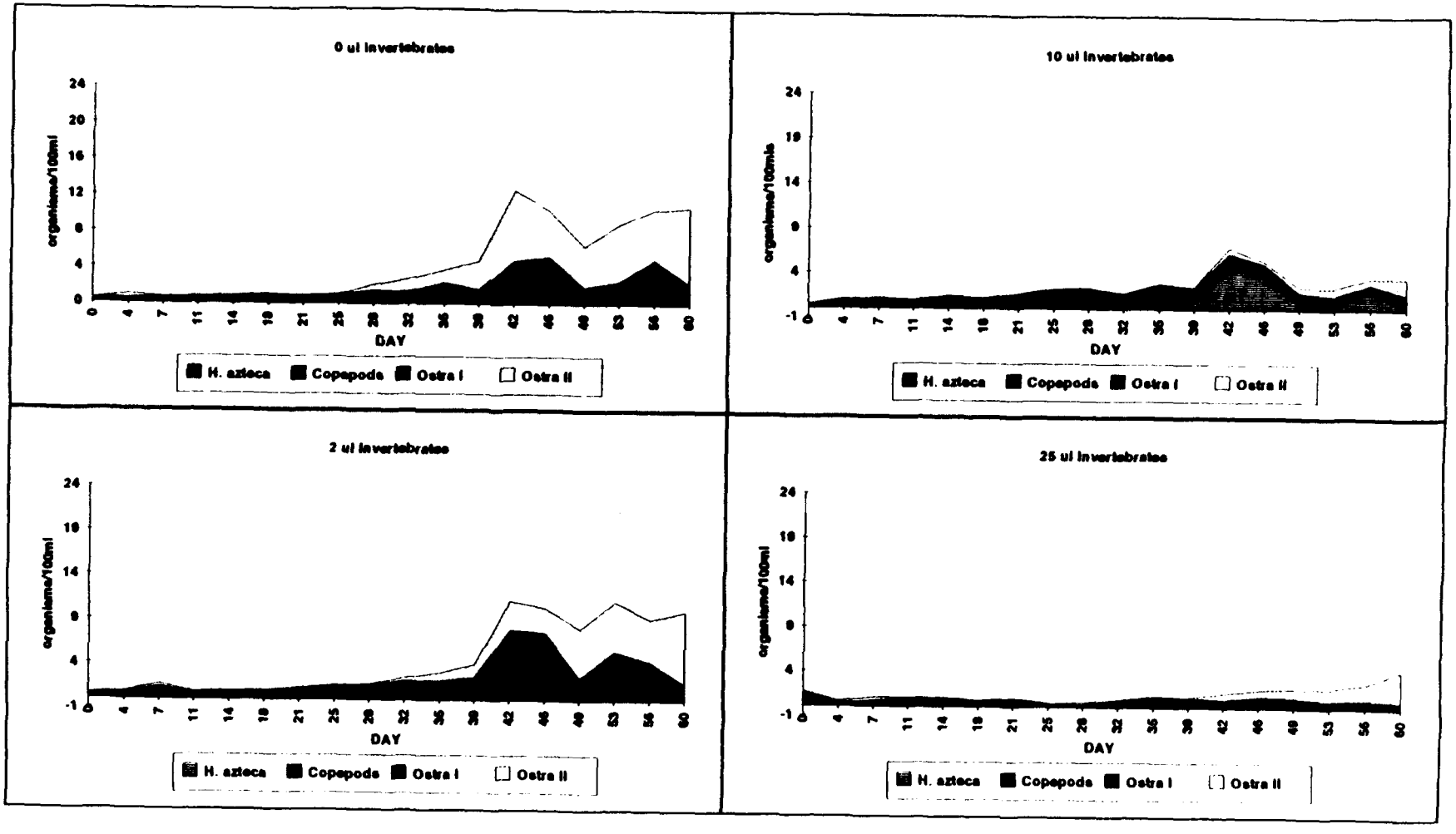


Fig. 22. Changes in invertebrate populations over time for all treatment groups. Numbers are treatment means and are useful for illustrating patterns in population dynamics.

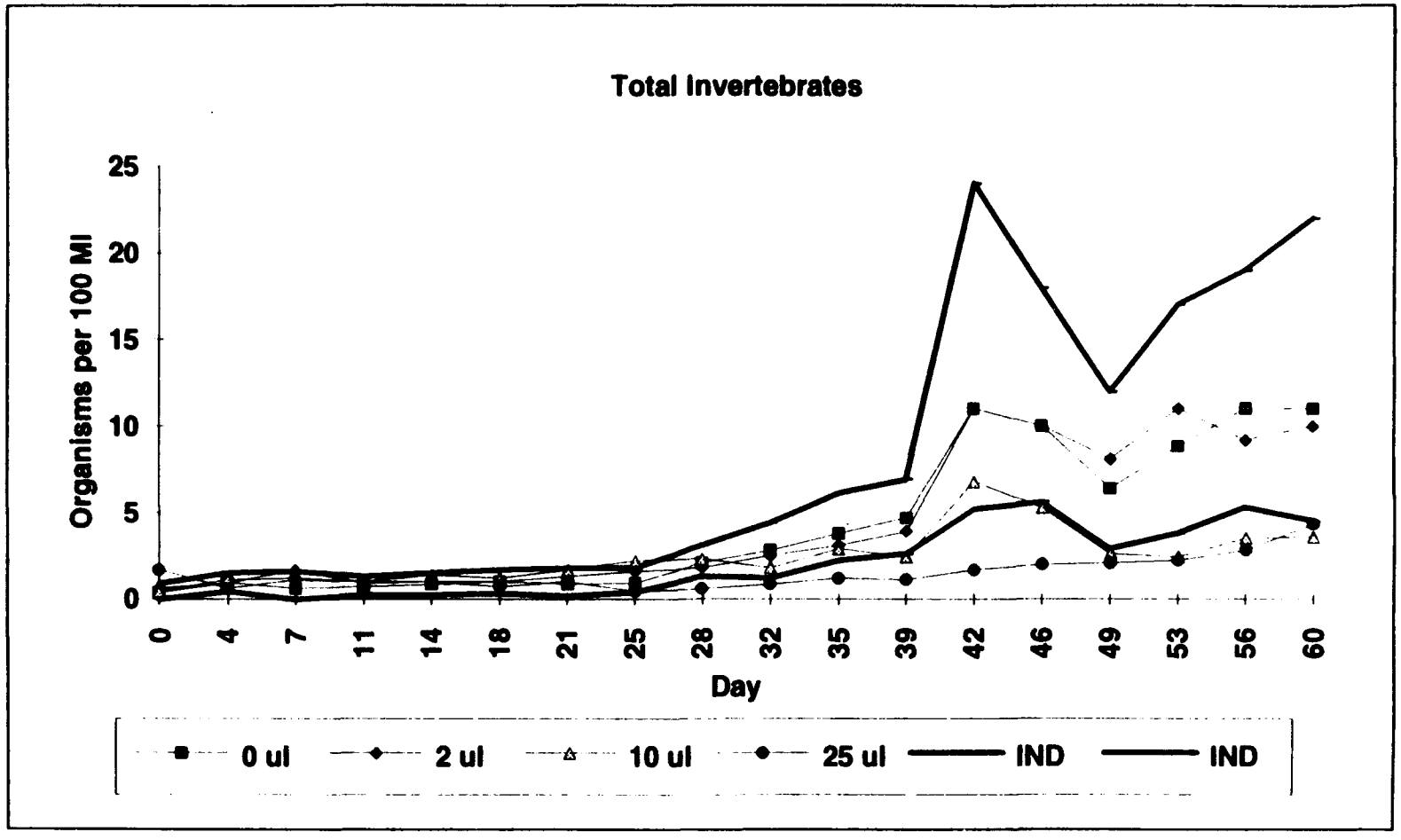


Fig. 23. Changes in total invertebrates over time. Values are treatment group means of organisms per 100 ml and are plotted with the Interval of Nonsignificant Difference (IND).

apparent recovery of the system from perturbation or stability of the 0 µl group (Fig. 23). An apparent bloom of *H. azteca and* ostracods (Ostracods II) occurred late in the experiment in most treatment groups and appeared to behave in a dose-response manner with decreasing total numbers in relation to dose (Figs. 24-25). These differences were detectably significant during the later bloom. Specifically, *H. azteca* populations remained relatively constant throughout the experiment with little variance due to very low population numbers until late in the test on days 39 - 60, when populations rose in all groups except the 25 µl group, where all the *H. azteca* were apparently dead. Ostracod II also contributed to the late bloom with the 0 and 2 µl groups having generally significantly greater numbers from day 28 on despite the large within group variance present (Fig. 25).

Multivariate Analysis

The significance levels for the three multivariate tests used are graphed in Figure 26. Euclidean distance was able to determine 12 significant days for the functional parameters and two for the structural parameters. Similarly, cosine of the vector and RIFFLE were able to determine significance 12 and six days for the functional parameters, and three and seven days for the structural parameters. Based on the fact that, for a test such as this with 18 consecutive sampling days, the probability for falsely rejecting the null hypothesis (a type I error) increases to 0.60 (with an α of 0.05 x 18 days = 0.60) if a rejection were to be based on a single significant observation. In view of this, significance, and a rejection of the null hypothesis, would occur with three observations of significant groupings of the data based on treatment effect (one erroneous significant effect could be expected due to chance every 10 sampling days-

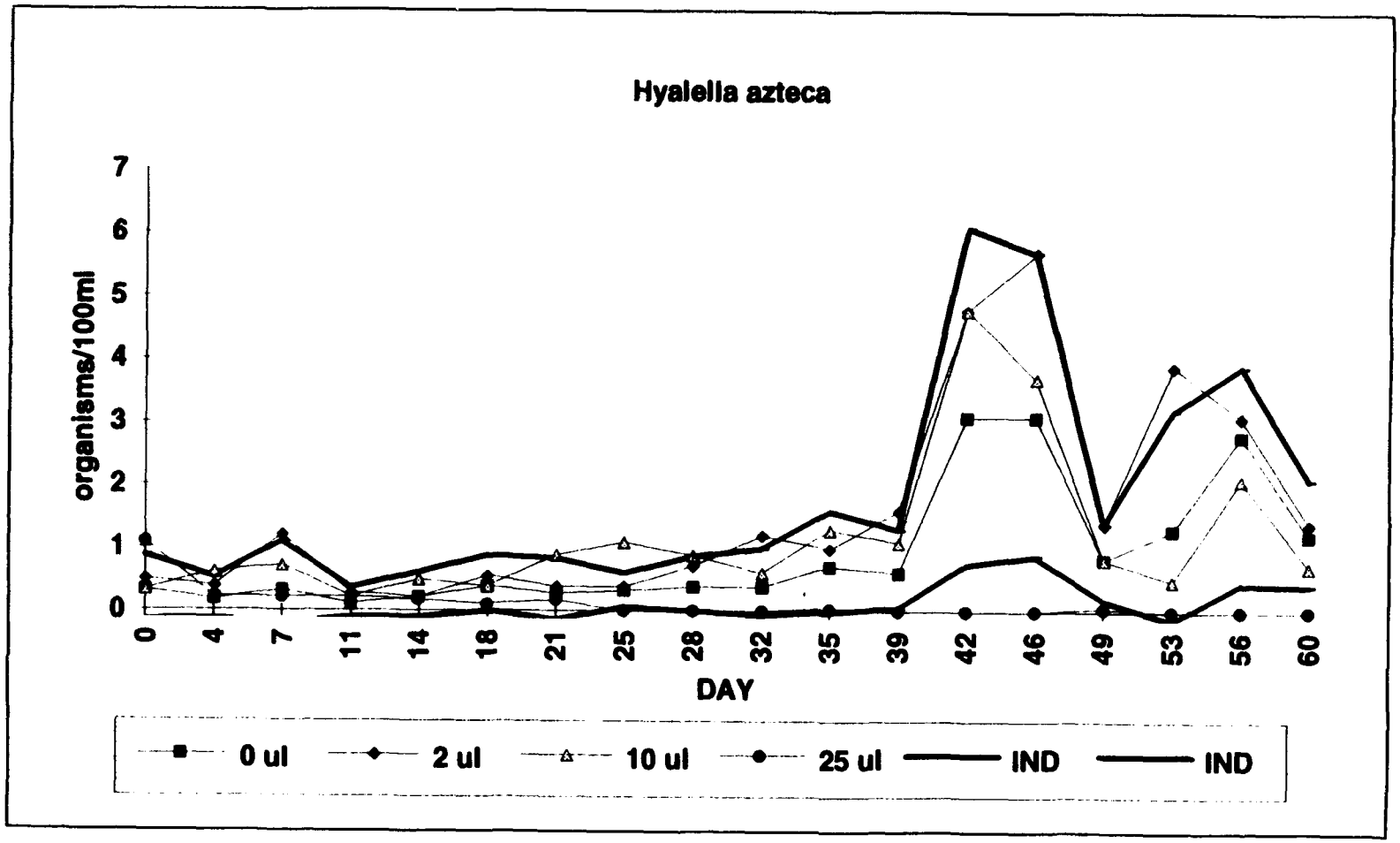


Fig. 24. Changes in *H. azteca* densities over time. Values are treatment group means of organisms per 100 ml and are plotted with the Interval of Nonsignificant Difference (IND).

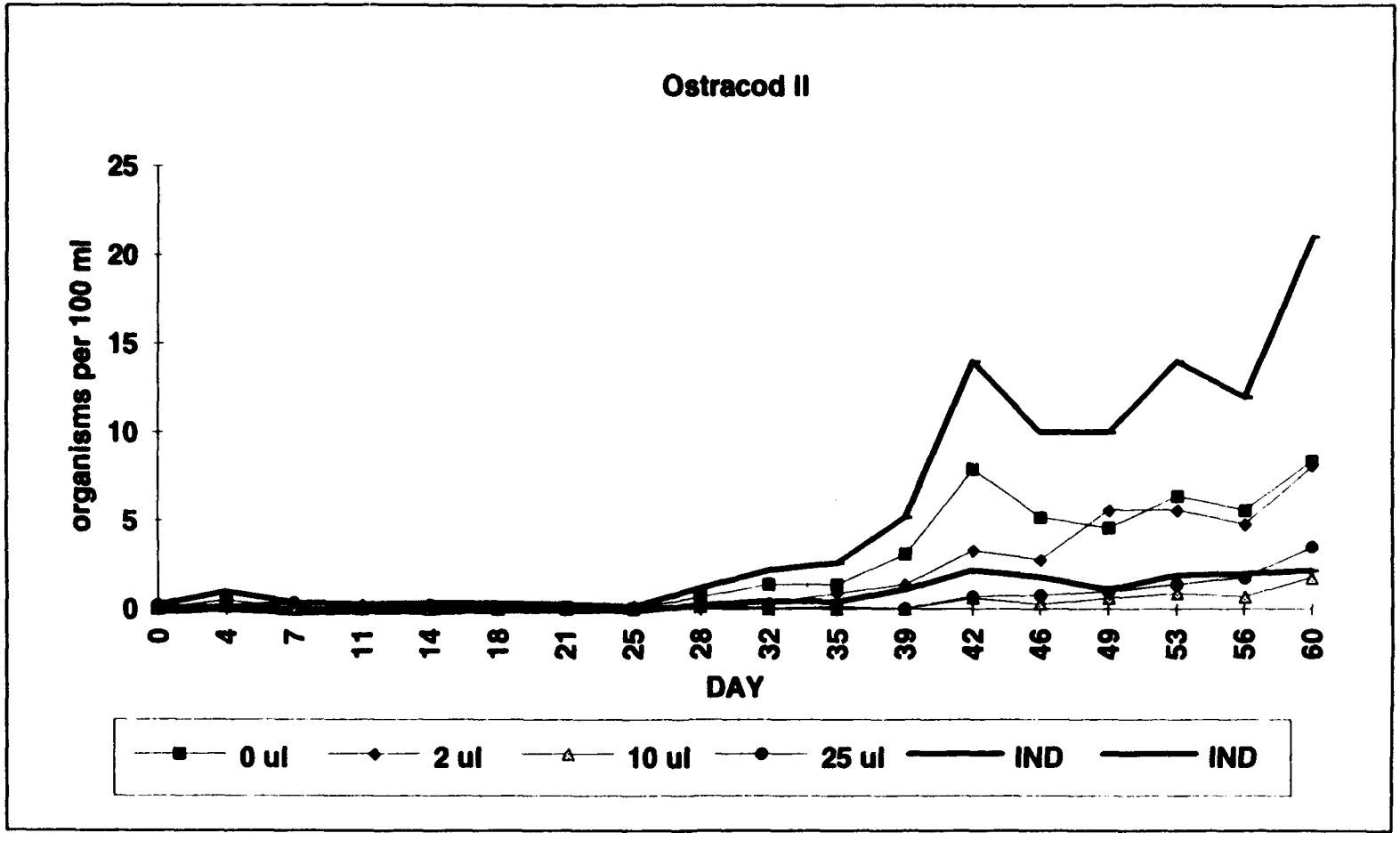


Fig. 25. Changes in Ostracod II densities over time. Values are treatment group means of organisms per 100 ml and are plotted with the Interval of Nonsignificant Difference (IND).

therefore, three significant days are reasonably required to demonstrate significance). Based on this, all three methods were able to demonstrate a significant treatment effect for both structural and functional parameters except Euclidean distance for the structural parameters (Fig. 26).

For functional parameters, cosine of the vector and Euclidean distance measures were able to pick out treatment effects much more frequently than Riffle. Oscillations in effect significance are apparent from the plots based on comparisons of all three multivariate tests and are similar to observations made by others with similar data sets (Landis et al., 1993a; Landis et al., 1993b; Landis et al., 1993c). Periods of no significant differences between treatment groups identified by all three multivariate tests may not be recovery of the dosed groups but merely periods of no detectable differences. The initial period of no significant differences between treatment groups identified by all three multivariate analyses (days 21-32) appears to correspond well with the similar dissolved oxygen levels observed during this period for all treatment groups (Figs. 5-7,10). Important variables identified with nonmetric clustering in determining treatment effect are shown in Tables 4 and 5. The overall importance of the pH and AMDO categories picked out by RIFFLE (Table 4) correspond well with observations made with the univariate results of increased respiration in the treated groups and this appears to change over time (Table 5). The slightly higher overall rankings of the AMDO parameters over the PMDO2 parameter may be due to the lower within group variances for the AMDO parameters generally observed throughout the test (Figs. 5-7).

Of the three methods used, nonmetric clustering (RIFFLE) appeared to do the best job of determining effect for the structural variables. Similar oscillations in effect significance are apparent from the plots based on comparisons of all three

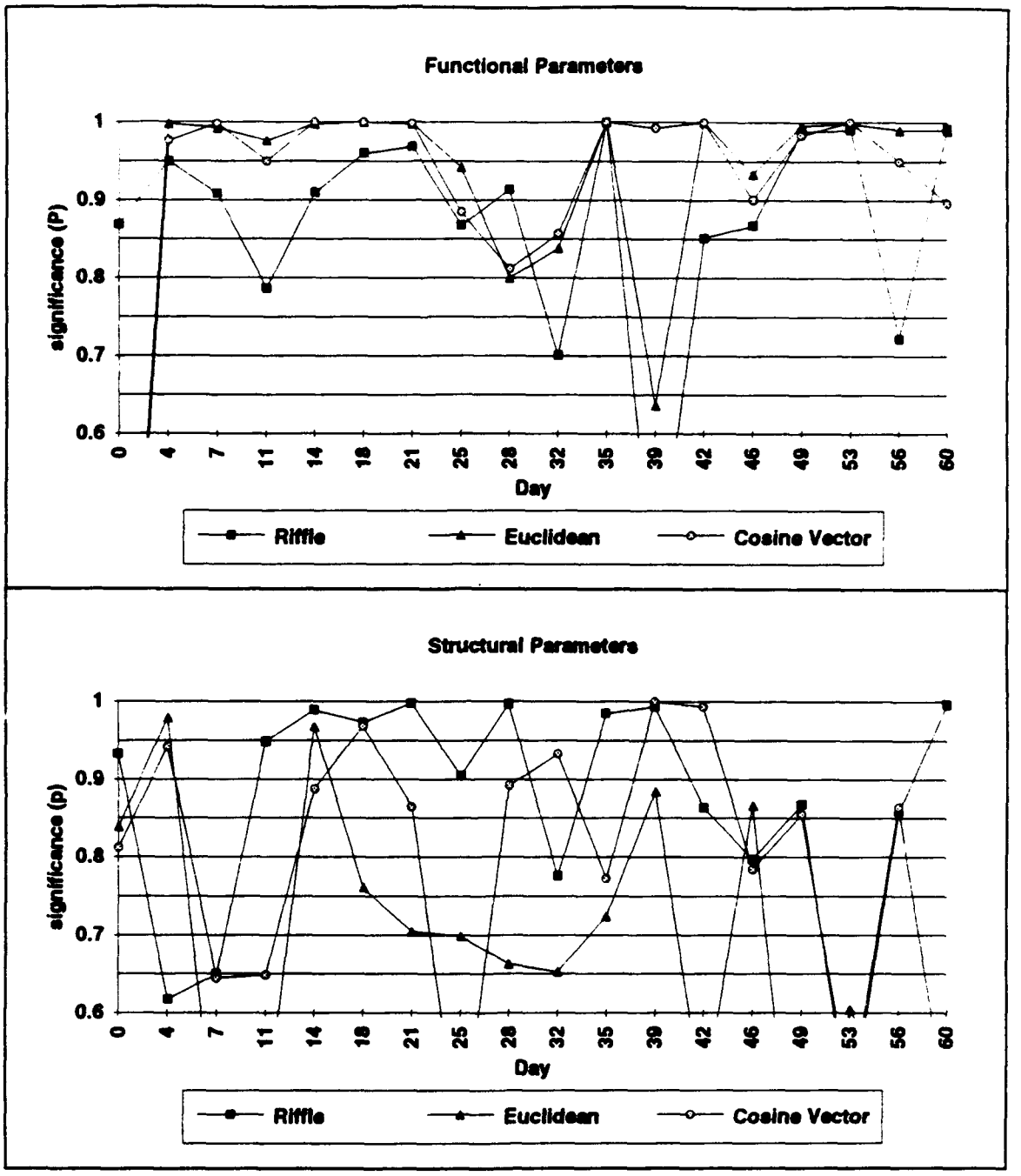


Fig. 26. Plots of significance over time of the Euclidean distance, cosine of the vector, and Riffle multivariate analyses. Critical values are 0.95 at the α=0.05 significance level.

Functional Variables	
Variable	Rank
pH	16
AMDO1	15
AMDO3	15
PMDO2	12
ABS	8
Structural Variables	
Scenedesmus	10
P. bursaria	9
Other Green Unicellular	8
Ciliates	7
Rotifers	7
H. azteca	7
Other Diatoms	6
Flagellates	5
Chlorella	4
Selenastrum	3
Lyngbya	3
Other Blue-Greens	3
Copepods	3
Ankistrodesmus	1
Nitzchia	1
Ostracods II	1

Table 4. Important variables according to success in determining treatment effect as determined by nonmetric clustering (RIFFLE). Values correspond to the frequency of obtaining a Proportional Reduction of Error (PRE) value greater than or equal to 0.5 throughout the entire test.

Day	Functional Parameters	Structural Parameters
0	AMDO1, **pH**, Abs, AMDO3	Rotifers, *Chlorella*-Other Diatoms, Other Green Unicellular, Ciliates
4	AMDO3, **pH**, AMDO1, PMDO2, Abs	*Chlorella*, Ciliates, Flagellates, ***P. bursaria***
7	**pH**, AMDO1, AMDO3, PMDO2	Other Diatoms, Ciliates, ***P. bursaria***
11	**pH**-PMDO2, AMDO1-Abs, AMDO3	Other Diatoms, Other Blue-greens, Flagellates, *Chlorella*
14	**pH**, AMDO1, PMDO2, AMDO3	*Lyngbya*-***P. bursaria***-*H. azteca*, ***Scenedesmus***, Copepods
18	PMDO2, **pH**, AMDO3, AMDO1	Flagellates, ***P. bursaria***, *H. azteca*, ***Scenedesmus***, Other Green Unicellular, Rotifers
21	AMDO1, **pH**, AMDO3, PMDO2	Rotifers, ***Scenedesmus***, Other diatoms, ***P. bursaria***
25	**pH**, AMDO1, PMDO2, Abs, AMDO3	***P. bursaria***, Flagellates, ***Scenedesmus***, Other Diatoms, *H. azteca*, Copepods
28	**pH**, AMDO3, AMDO1	***P. bursaria***, *H. azteca*, ***Scenedesmus***, *Lyngbya-Chlorella*, Rotifers
32	AMDO1, AMDO3, PMDO2	Ciliates, Rotifers-*Lyngbya*, Other Green Unicellular
35	**pH**, PMDO2, Abs	*Ankistrodesmus*, ***Scenedesmus***-Other Diatoms-Ciliates, Other Blue-Greens-Ostracod II
39	AMDO3, PMDO2, AMDO1, Abs	***P. bursaria***, Other Green Unicellular, *H. azteca*, Other Diatoms
42	PMDO2, AMDO3, AMDO1, **pH**	Other Green Unicellular, Ciliates, *H. azteca*, ***Scenedesmus***
46	AMDO3, AMDO1, **pH**	***P. bursaria***, Copepods, Ciliates, *Selenastrum*
49	Abs, **pH**, AMDO1	***Scenedesmus***, Flagellates, Rotifers, *Nitzchia*
53	**pH**-AMDO3, AMDO1	***P. bursaria***, *Selenastrum*, Other Green Unicellular, Other Blue-Greens
56	AMDO3, Abs, PMDO2-**pH**	***Scenedesmus***, *H. azteca*, Other Green Unicellular
60	**pH**, AMDO1, AMDO3	Ciliates, Rotifers, *Selenastrum*

Table 5. Important variables ranked according to contribution for each sampling day as determined by nonmetric clustering for functional and structural parameters. Highlighted parameters were important throughout the test. Note: hyphen between values denotes equal rank.

multivariate tests (Fig. 26). The long period of no observed significant differences between treatment groups identified by all three multivariate analyses (days 46-56) is apparently due to the relative similarity of the major structural groups during this period (Figs. 14, 19, and 23). Important variables identified by RIFFLE in determining treatment effect are also shown in Tables 4 and 5. These results also correspond well with the results of the univariate methods and support observations of the overall importance of certain structural groups in determining treatment effect due to differences in numbers. A similar change in importance over the course of the test is also evident.

Bacteria

The results of the comparisons of the bacterial cell counts were inconclusive due to the apparent nonlinear relationships of the data and the lack of replication precluding the use of ANOVA. Positive Pearson correlation's in bacterial numbers were found for comparisons of the 0, 2, and 10 µl treatment groups with time, and between the 0 and 2 µl treatment groups (Table 6).

	Day	0 µl	2 µl	10 µl
0 µl	**0.59**			
2 µl	**0.61**	**0.72**		
10 µl	**0.51**	0.27	0.32	
25 µl	0.19	0.18	0.45	0.07

Table 6. Pearson correlation coefficients of bacterial cells per 15 fields obtained from direct counts. Bold values are significant. $R_{crit.\ 0.05\ (2),\ 16} = 0.468$. H_0: $\rho=0$

Subsequent linear regression analysis of all treatment groups with time revealed significant relationships and positive slopes for the 0, 2, and 10 µl treatment groups with intercepts not significantly different from zero at the $\alpha = 0.05$

significance level (Table 7). However, examination of line fit plots revealed high values for all three of these treatment groups (Fig. 27). Due to the known potential for linear regression to be overly influenced by these large values, this precluded any comparison of slopes for dose-response relationships. From Figure 27, the relationship between the 0 and 2 µl groups is apparent and the 0, 2, and 10 µl groups appear to have a nonlinear relationship with time. The 25 µl group appears to have been depressed in growth over time, possibly due to a toxic effect of the Jet-A resulting in the absence of a significant relationship with time. However, the lack of replication precluding the use of ANOVA or the Interval of Nonsignificant Difference (IND), and the apparent nonlinear relationships of the treatment groups with time prevents the use of this data for drawing any further conclusions.

	0 µl	2 µl	10 µl	25 µl
R	**0.59**	**0.61**	**0.51**	0.19
Fcalc	**8.56**	**9.66**	**5.77**	0.58
p value	**0.01**	**0.01**	**0.03**	0.46
Regression Coefficients				
Intercept	125	109	179	**253**
t statistic	1.34	1.27	1.49	**3.93**
p value	0.2	0.22	0.15	**0**
Slope	**7.77**	**7.59**	**8.22**	1.4
t statistic	**2.93**	**3.11**	**2.4**	0.76
p value	**0.01**	**0.01**	**0.03**	0.46

Table 7. Results of linear regression analysis of bacterial cells per 15 fields over time. Bold values are significant. For R, $R_{crit\ 0.05\ (2),16} = 0.468$; H_0: $\rho = 0$. For the F statistic; H_0: all coefficients are zero. For the t statistics; H_0: coefficient is zero.

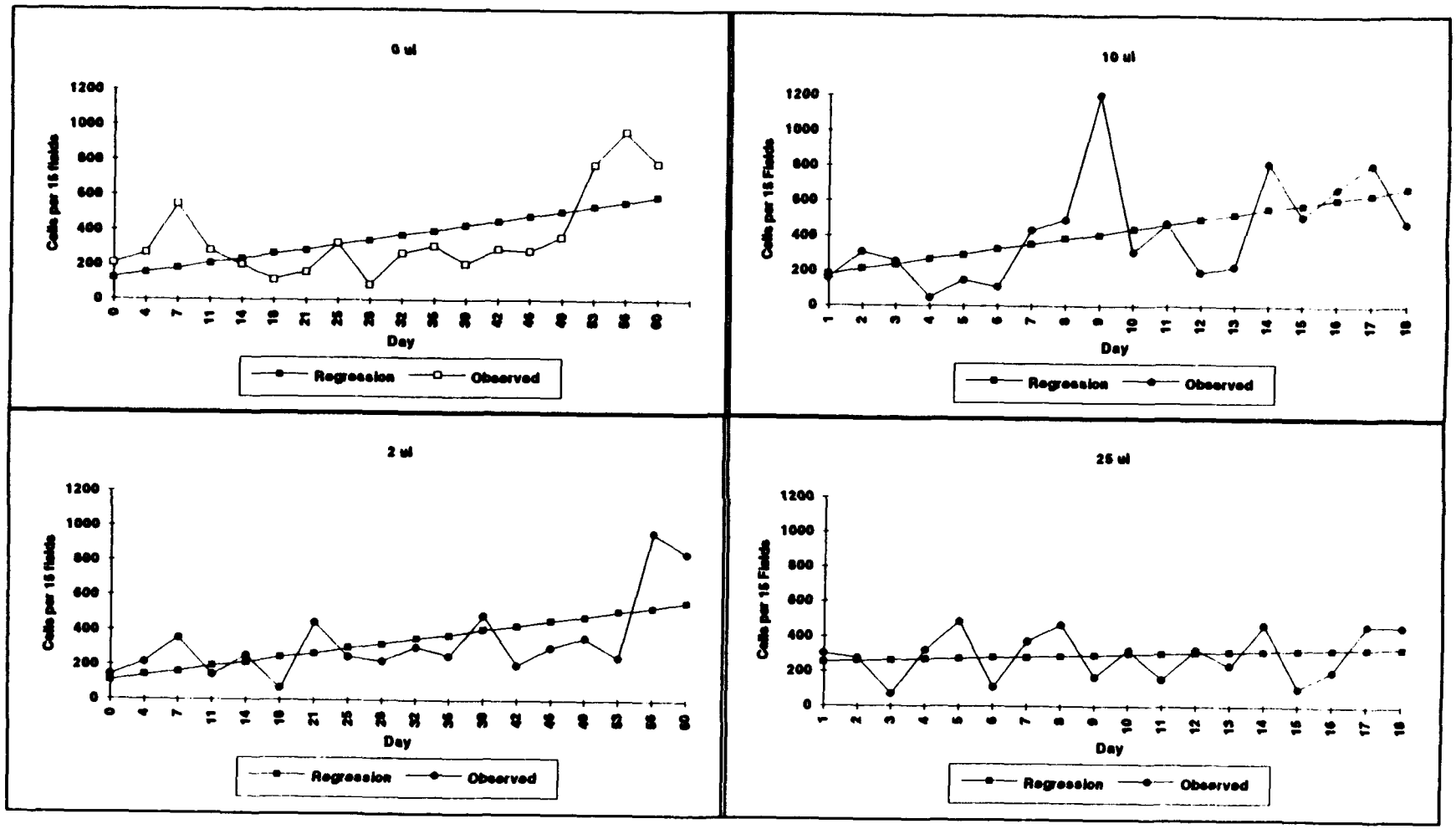


Fig. 27. Regression vs. observed data plots of bacterial cells per 15 fields generated by direct count techniques for all treatment groups.

Total Organic Carbon

Total organic carbon analysis yielded significantly increasing average amounts of organic carbon in relation to treatment (Table 8) although the uncertainty as to its origin limited the usefulness of the data. Based on ANOVA, a significant difference was observed for all treatment groups. The progressively increasing amounts with treatment may have been due to remaining hydrocarbons in the sediment, or to the presence of degradative microbes utilizing the Jet-A as an energy source in the sediment layer.

Treatment groups	0 µl	2 µl	10 µl	25 µl
Treatment means	**0.051**	**0.059**	**0.067**	**0.083**
ANOVA TABLE				
Source	Degrees of freedom	Sum Squares	Mean Square	F
Between	3	0.0034	0.0011	**11.000**
Within	19	0.0018	0.0001	
Total	22	0.0052		
Critical F = 3.13 (0.05,3,19)	Decision: reject H_0: all groups equal			

Table 8. Treatment means and ANOVA results for total organic carbon (TOC). Mean values are in percent carbon. Treatment groups were significantly different based on ANOVA.

DISCUSSION

The results of this research indicate that the technique of incorporating contaminated sediment produced a significant response in the MFC multispecies test and may be a useful technique for the exposure and evaluation of other sediment contaminants in a multispecies test system. Moreover, the results indicate that the cross inoculation procedures designed to ensure replicability as well as the size of the test chambers, may be inadequate. Finally, the results of this MFC multispecies test are not incompatible with observations of others who question the existence of stability in biological communities.

The response of the MFC to the treatment indicated that an initial universal depression in dissolved oxygen levels occurred due to an increase in respiration, which was followed by an algal bloom of primarily blue-green algae, and a general depression in protozoa and invertebrate populations later in the test. The initial functional response of increased respiration appeared to be due to the transfer perturbation increasing the availability of sediment born heterotrophic substrates since the untreated group was included in the initial response. However, an increase in respiration is also a suspected response to toxic stress (Odum, 1985) and this could explain the later differences observed between the reference and treated groups. Also, although the treatment amounts were small, the added substrate provided by the complex hydrocarbons present in the treated groups may have been a factor in the observed increase in respiration. Hydrocarbons are known to be degraded in aquatic systems by microbes under suitable conditions, the rate of which varies with mixture. Some of this degradation is known to proceed best when associated with sediments and can be influenced by the availability of nutrients (Rheinheimer, 1985). Since the Jet-A was associated with the sediment in this test and the availability of nutrients in

the media was high at the outset, it seems plausible that this could have been responsible for some of the increase in respiration. Most likely however, the observed dynamics are due to a combination of the above factors.

The responses of the structural parameters evaluated in this test are similar to those typically observed in these tests. Towards the end of the test period, the observed bloom of blue -green algae is a common response when the xenobiotic is not selectively toxic and nutrient limitation occurs, favoring species capable of producing nutrients. The general depression in numbers of the protozoa and invertebrate groups is also a common response and is assumed to have been due to direct or indirect toxic effects of the Jet-A.

One interesting exception to this is the response of *P. bursaria* in this test. Throughout much of the test, *P. bursaria* numbers were generally four times as dense in the 25 µl treatment group as in the reference 0 µl group. The 2 and 10 µl groups also appeared to have larger numbers although this was not generally statistically significant based on the IND. Since *P. bursaria* are known opportunists capable of both autotrophic and heterotrophic activity, and are limited in specific competitive ability (Landis, 1988), these dynamics may indicate that competition for some resources was negligible in these groups. Theoretically, a species with high genetic variability and low competitive ability could survive the impact of the xenobiotics and take advantage of the subsequent lack of competition for resources. However, not enough is known about the ecology of these organisms to provide a definite answer.

Due to the large pore capacity of the overlying silica sand MFC sediment allowing access of the overlying water and detritus to the spiked layer, and the incorporation of the powdered cellulose and chitin in the spiked sediment layer providing sorptive substrate for the xenobiotics, an effective simulation of natural freshwater sediment contamination from underlying sources was also achieved.

This technique may also be valid for use with natural contaminated freshwater sediment either through the use of sediment dilution to obtain concentration-effect information (Giesy et al., 1990; Landrum et al., 1990), or simply with whole sediments from various contaminated sites. Problems with obtaining acceptable control and reference sediments, as well as with the dilution method itself, would have to be addressed. A suitable "clean" non contaminated reference site sediment would have to be obtained with physiochemical properties which bracket that of the contaminated sediment (ASTM E 1391-90, 1991). In addition, a control sediment (Adams et al., 1985), either natural or artificially prepared, would need to be obtained with a known composition and quality for which baseline information is available that demonstrates no toxicity (ASTM E 1391-90, 1991).

Dilution methods would also have to be acceptable. Currently, there is little information available on the most appropriate methods for diluting test sediments to obtain graded contaminant concentrations or concerning the methodological effects of such dilutions (Burton, 1991). Materials used for the dilution of the contaminated sediment must also be uncontaminated and have physiochemical properties similar to the contaminated sediment so as not to affect the contact time of the interstitial water with the contaminated sediment. Also, the effect of disrupting the sediment integrity during sampling and manipulation must be considered.

The results of this MFC multispecies microcosm test also indicate that the cross inoculation procedure specified in the protocol and used in this research is inadequate. The occurrence of differences in initial numbers of virtually all structural components indicates that assumptions and procedures to ensure replicability were not adequate. This may have been due to inadequate volume transfer during the cross inoculation procedure, resulting in small initial

differences in species abundances between replicates. This could have changed community interactions significantly enough to account for the differences in abundances between replicates during the three day interval between the last cross inoculation and the initial counts. Algal and protozoan populations could theoretically have increased significantly during this three day interval. Observations of differences in initial numbers of some of the invertebrate groups, though noticeable different, can be discounted due to low initial numbers. These initial differences amount to essentially one total organism between treatment groups which could have been purely a chance event. Due to these observations, I would recommend both increased volume during the cross inoculation from 10 percent of volume to 20 - 40 %, as well as a cross inoculation just preceding sampling and dosing on day zero to improve the replicability of this particular protocol.

Perhaps most interestingly however, is the observation that the test appears to have been inadequate on both spatial and temporal scales. Spatially, this is very evident since, again, some of the invertebrate groups had very low initial population numbers. Differences of essentially portions of one individual between treatment groups were observed and treatment groups were being compared on the basis of a very few total individuals. The genomic makeup of these individuals may, or may not, be representative of the worlds population of these particular organisms. In any event, if these individuals were only present in some replicates and were to vary considerably in their sensitivities to the xenobiotics, or were simply to die of natural causes, the results of the multispecies test could have become overly dependent on the simple presence or genetic makeup of these few individuals due to their importance in their respective communities. The resulting changes in community structure and function would not be representative of a natural ecosystem where a larger

genome of the particular species in question would be present. Any multispecies test system must be sized appropriately to avoid these problems and must be at least large enough to include a representative number of all potentially included more common organisms identified with species abundance distributions. In my view, this size must be based on the largest of these species in order to include a representative sample of all common species populations.

Temporally, virtually all of the structural variables monitored with the univariate analyses indicate that the treated groups had not returned to a pre exposure or control state upon the completion of the test based on the IND, or that a stability of the 0 µl control group was demonstrated. Landis et al. (1993b), have observed, in similar tests with similar oscillations in effect significance using multivariate analyses, changes in multidimensional representations of treatment groups which suggest that a return to a control, or pre exposure state, may be impossible. Periods of apparent recovery may simply be illusions created by the reduction in dimensionality that accompanies the usual two dimensional representations of the data. Questions have been raised (Landis et al., 1993d) regarding the apparent nonlinear nature of these oscillations in relation to complex systems, chaotic dynamics, the importance of historical events, and the irreversibility of these systems. These observations imply that the generic vs. specific argument concerning these multispecies tests, and the reference to a control, or stable, state may be invalid. It was suggested that possibly a more workable solution or definition of recovery may be the ability to distinguish treatment effects from historical stress, whether these are due to prior pollutants, or natural factors. Indeed, as Connell and Sousa (1983) have suggested, in an extensive review of the existence of stability in the classic sense, that stability may not exist and that perhaps a more workable definition would simply be persistence of a given ecosystem within bounds.

In any event, In my view, the fact remains that the observed nonlinear dynamics, apparent in the above mentioned multidimensional projections, may still simply be due, in part at least, to founder effect. The small populations present in these microcosm multispecies tests are, including the non treated groups, restrictions of a genome, which is in itself a treatment effect. Connell and Sousa (1983) have suggested that the bounds of population fluctuation may narrow as the spatial scale becomes larger. Clearly, these dynamics need to be investigated in further detail on a larger spatial and temporal scale in a natural setting, where the test system is really a part of a much larger system in which the effects of natural immigration from unexposed sources and the incomplete dispersal of xenobiotics is the norm and physical dynamics are uncontrolled.

Recently, the EPA has removed requirements, in most cases, for acute effects of pesticides on birds, and outdoor mesocosm studies (Fisher, 1992). The justification is based on the fact that, for making risk decisions, the information supplied by these studies does not contribute much beyond that supplied by lower tiered, less expensive tests. Also, the information generated by these studies takes substantial time for review by the EPA staff. The intended result of these changes being that risk assessment will be enhanced due to using resources reviewing more readily available lower tiered information and incident reports, resulting in risk management in a more timely manner than has been previously achieved.

Historically, mesocosm research has been hampered by failing to demonstrate ecological damage due to cost factors usually limiting replicates to three, which usually have high variability. Also, questions concerning the degree to which effects can be assigned to the test chemical and not some outside factor, the difference between mesocosm studies, and extrapolation of results to more realistic aquatic ecosystems have not been adequately answered.

These studies should be continued in view of the above mentioned new research developments. Answers to most of the problems associated with these mesocosm studies may be become more apparent with new developments in statistical and multidimensional visualization techniques providing new methods for the definitive evaluation of these types of studies, as well as previously impossible insights into the validity of previously established theoretical ecological paradigms. These new techniques may be able to see through the uncontrolled previous and concurrent natural and anthropogenic stressors causing high variability and tease out the effects of the xenobiotic in question. However, if nonlinear and chaotic dynamics, irreversibility, and the importance of small initial differences emerge as properties of these systems, the extrapolation of specific results to other natural systems may be impossible. These types of studies may lead to the development of new standards for the protection and management of ecosystems as persistence within bounds under constant natural and anthropogenic stressors.

CONCLUSIONS

1. The sediment spiked Jet-A turbine fuel produced statistically significant functional and structural responses in the MFC multispecies test observed through the use of both univariate and multivariate statistical techniques.

2. The method of incorporating spiked sediment into an established multispecies test system is a useful technique and may merit further study using other types of test material, including contaminated natural sediment.

3. The cross inoculation procedure as specified in the MFC protocol is inadequate.

4. The MFC multispecies test needs to increase on a spatial scale sufficient to include representative numbers of all individual species populations.

5. The observed instability of the control group and failure of the treated groups to return to a pre exposure state based on structural data are not incompatible with the observations of others questioning the existence of stability.

REFERENCES

Adams, W.J., Kimerle, R.A., and Mosher, R.G. 1985. Aquatic safety assessment of chemicals sorbed to sediments. In R.D. Caldwell, R. Purdy, and R.C. Bahner, eds., *Aquatic Toxicology and Hazard Assessment: Seventh Symposium. ASTM STP 854.* American Society for Testing and Materials. Philadelphia. pp.429-453.

Aelion, C.M., and Bradley, P.M. 1991. Aerobic biodegradation potential of subsurface microorganisms from a jet fuel contaminated aquifer. *Appl. and Environ. Microbiol.* **57**(1):57-63.

ASTM D 439-89. 1990. Standard specification for automotive gasoline. vol. 05.01. American Society for Testing and Materials, Philadelphia. pp. 156-169.

ASTM D 1655-89. 1990. Standard specification for aviation turbine fuels. vol. 05.01. American Society for Testing and Materials, Philadelphia. pp.619-627.

ASTM D 2699-88a. 1990. Test method for knock characteristics of motor fuels by the research method. vol. 05.04. American Society for Testing and Materials, Philadelphia. pp. 12-25.

ASTM D 2700-88a. 1990. Test method for knock characteristics of motor and aviation fuels by the motor method. vol. 05.04. American Society for Testing and Materials, Philadelphia. pp. 26-46.

ASTM E 1366-91. 1991. Standard practice for the standardized aquatic microcosm: freshwater, vol 11.04. American Society for Testing and Materials, Philadelphia. pp. 1017-1051.

ASTM E 1383-90. 1991.Standard guide for conducting acute sediment freshwater toxicity tests with freshwater invertebrates, vol 11.04. American Society for Testing and Materials, Philadelphia. pp. 1017-1051.

ASTM E 1391-90. 1991.Standard guide for the collection, storage, characterization, and manipulation of sediments for toxicological testing, vol 11.04. American Society for Testing and Materials, Philadelphia pp. 1017-1051.

Bedford, B. 1984. The role of ecosystem science in USEPA's mission to regulate toxic substances. *Environ. Manage.* **8**:377-383.

Burton, G.A., Jr. 1991. Assessing the toxicity of freshwater sediments. *Environ. Toxicol. Chem.* **10**:1585-1627.

Burton, G.A., Jr., Lazorchak, J.M., Waller, W.T., and Lanza, G.R. 1987. Arsenic toxicity changes in the presence of sediment. *Bull. Environ. Contam. Toxicol.* **38**:491-499.

Cairns, J., Jr., 1983. The case for simultaneous toxicity testing at different levels of biological organization. In W.E. Bishop, R.D. Cardwell, and B.B. Heidolph, eds., *Aquatic Toxicology and Hazard Assessment: Sixth Symposium, ASTM STP 802.* American Society for Testing and Materials, Philadelphia. pp. 1017-1051.

Cairns, J., Jr. 1984. Multispecies toxicity testing. *Environ. Toxicol. Chem.* **3**:1-3.

Cairns, J., Jr. 1986. The myth of the most sensitive species. *BioScience.* **36**:670-672.

Cairns, J., Jr. 1993. The threshold problem in toxicology. *Ecotoxicology.* **1**:3-16.

Cairns, J., Jr., and Pratt, J.R. 1986. On the relation between structural and functional analyses of ecosystems. *Environ. Toxicol. Chem.* **5**:785-786.

Carpenter, S.R., Kitchell, J.F., and Hodgsdon, J.R. 1985. Cascading trophic interactions and lake productivity. *Bioscience.* **35**:634-639.

Chapman, P.M. 1989. Current approaches to developing sediment quality criteria. *Environ. Toxicol. Chem.* **8**:589-599.

Coleman, A.W. 1980. Enhanced detection of bacteria in natural environments by fluorochrome staining of DNA. *Limnol. Oceanogr.* **25**(5): 948-951.

Connell, J.H., and Sousa, W.P. 1983. On the evidence needed to judge ecological stability or persistence. *The American Naturalist.* **121**(6):789-824.

Conquest, L.L., and Taub, F.B. 1989. Repeatability and reproducibility of the standard aquatic microcosm: statistical properties. In U.M. Cowgill and L.R. Williams. eds., *Aquatic Toxicology and Hazard Assessment: 12th Volume, ASTM STP 1027.* American Society for Testing and Materials, Philadelphia. pp. 159-177.

Fienberg, S.E. 1985. *The analysis of cross-classified categorical data.* Mit Press, Cambridge, Ma.

Fisher, L. 1992. Memorandum: *Decisions on the ecological, fate, and effects task force.* Office of Pesticides and Toxic Substances, U.S. Environmental Protection Agency.

Fransisco, D.E., Mah, R.A., and Rabin, A.C. 1973. Acridine orange-epiflourescence technique for counting bacteria in natural waters. *Trans. Am. Microsc. Soc.* **92**:416-421.

Giddings, J.M. 1981. Laboratory tests for chemical effects on aquatic population interactions and ecosystem properties. In A.S. Hammons, ed., *Methods for Ecological Toxicology, A Critical Review of Laboratory Mulispecies Tests.* Ann Arbor Science Publishers Inc., Ann Arbor. pp. 23-91.

Giesy, J.P., Rosiu, C.J., Graney, R.L., and Henry, M.G. 1990. Benthic invertebrate bioassays with toxic sediment and pore water. *Environ. Toxicol. Chem.* **9**:233-248.

Good, I.J. 1982. An index of separateness of clusters and a permutation test for its significance. *J. Statist. Comp. Simul.* **15**: 81-84.

Guiney, P.D., Sykora, J.L., and Keleti, G. 1987. Qualitative and quantitative analyses of petroleum hydrocarbon concentrations in a trout stream contaminated by an aviation kerosene spill. *Environ.Toxicol. Chem.* **6**: 105-114.

Heath, R.T. 1980. Are microcosms useful for ecosystem analysis? In J.P. Giesy ed., *Microcosms in ecological research.* Springfield , VA: Technical Information Center. pp. 33-347.

Hutchins, S.R., Downs, W.C., Wilson, J.T., Smith, G.B., Kovacs, D.A., Fine, D.D., Douglass, R.H. and Hendrix, D.J. 1991. Effect of nitrate addition on biorestoration of fuel contaminated aquifer: field demonstration. *Groundwater.* **20**(4):571-580.

Ingersoll, C.G., and Nelson, M.K. 1990. Testing sediment toxicity with *Hyalella azteca* (Amphipoda) and *Chironomus riparius* (Diptera). In W.G. Landis and W.H. van der Schalie, eds., *AquaticToxicology and Risk assessment: Thirteenth Volume, ASTM STP 1096.* American Society for Testing and Materials, Philadelphia. pp. 93-109.

Johnson, A.R. 1988a. Diagnostic variables as predictors of ecological risk. *Environ. Manage.* 12(4): 515-523.

Johnson, A.R. 1988b. Evaluating ecosystem response to toxicant stress: a state space approach. In W.J. Adams, G.A. Chapman, and W.G. Landis, eds., *Aquatic toxicology and Hazard Assessment: 10th Volume. ASTM STP*

971. American Society for Testing and Materials. Philadelphia. pp. 275-285.

Kersting, K. 1988. Normalized ecosystem strain in mico-ecosystems using different sets of state variables. *Verh. Internat. Verein. Limnol.* **23**:1641-1646.

Kimball, K.D., and Levin, S.A., 1985. Limitations of bioassays: The need for ecosystem-level testing. *Bioscience.* **35**:165-171.

Landis, W.G. 1988. Ecology. In H. -D. Gortz, ed., *Paramecium*. Springer-Verlag, New York. pp. 421-436.

Landis, W.G., Matthews, R.A., Markiewicz A.J., and Matthews, G.B. 1993a. Multivariate analysis of the impacts of the turbine fuel JP-4 in a microcosm toxicity test with implications for the evaluation of ecosystem dynamics and risk assessment. *Ecotoxicology*. In Press.

Landis, W.G., Matthews, R.A., Markiewicz A.J., and Matthews, G.B. 1993b. Nonlinear oscillations detected by multivariate analysis in microcosm toxicity tests with complex toxicants: Implications for biomonitoring and risk assessment. ASTM. In Press.

Landis, W.G., Matthews, R.A., Markiewicz, A.J., Shough, N.J. and Matthews, G.B. 1993c. Multivariate analysis of the impacts of Jet-A turbine fuel using a standard aquatic microcosm toxicity test. *J. Env. Sci.* In Press.

Landis, W.G., Matthews, R.A., and Matthews, G.B. 1993d. Memo to Jeffrey M. Giddings, Editor, SETAC News. 10 August, 1993.

Landrum, P.F., Tsymbal, V.N., Nelson, M.K., Ingersoll, D.C., Gossiax, D.C., Burton, G.A., and Sasson-Brickson, G. 1990. Sediment associated contaminant toxicity assessment by dilution experiments. *Abstracts*, 11th annual meeting, Society of Environmental Toxicology and Chemistry, Arlington, VA. November 11-15, p. 107.

Lee, G.F., and Jones, R.A. 1984. Water quality significance of contaminants associated with sediments: An overview. In K.L. Dickson, A.W. Maki, and W.A. Brungs, eds., *Fate and Effects of Sediment-Bound Chemicals in Aquatic Systems*. Pergamon Press. New York. pp. 1-34.

Leffler, J.W. 1984. The use of self-selected, generic aquatic microcosms for pollution effects assessment. In H.H. Harris, ed., *Concepts in Marine Pollution Measurements*. Maryland Sea Grant College, Univ. Maryland, College Park, Maryland.

Lyman, W.J. 1984. In K.L. Dickson, A.W. Maki and W.A. Brungs, eds., *Fate and Effects of Sediment-Bound Chemicals in Aquatic Systems.* Pergamon Press. New York. pp. 378-387.

Madsen, E.L., Sinclair, J.L., and Ghiorse, W.C. 1991. In situ biodegradation: microbiological patterns in a contaminated aquifer. *Science.* **252**:830-833.

Matthews, G. B., and Hearne, J.W. 1991. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **13**(2):175-184.

Matthews, G.B., Matthews, R.A., and Hachmoller, B. 1991a. Mathematical analysis of temporal and spatial trends in the benthic macroinvertebrate communities of a small stream. *Can. J. Fish. Aquat. Sci.* **48**(11):2184-2190.

Matthews, R.A., Matthews, G.B., and Ehinger, W.J. 1991b. Classification and ordination of limnological data: a comparison of analytical tools. *Ecological modelling.* **53**:167-187.

McMahon, J.A., Phillips, D.L., Robinson, J.V., and Schimpf, D.J. 1978. Levels of biological organization: an organisms-centered approach. *BioScience.* **28**:700-704.

Meyer-Reil, L.-A. 1987. Seasonal and spatial distribution of extracellular enzymatic activities and microbial information of dissolved organism substrates in marine sediments. *Appli. Environ. Microbiol.* **53**:1748-1755.

Moriarty, F. 1988. *Ecotoxicology: the study of pollutants in ecosystems.* Second edition. Academic Press, San Diego, CA, USA.

Mount, D.L. 1985. Scientific problems in using multispecies toxicity tests for regulatory purposes. In J. Cairns Jr., ed., *MultispeciesToxicity Testing.* Pergamon Press, New York. pp. 13-18.

Noreen, E.W. 1989. *Computer Intensive Methods for Testing Hypotheses.* Wiley- Interscience, New York, NY.

Odum, E.P. 1985. Trends expected in stressed ecosystems. *Bioscience.* **35**:419-422.

Odum, E.P. 1990. Field experimental tests of ecosystem level hypotheses. *Trends Ecol. Evol.* **5**:203-236.

Oust. 1990. *LUST Trust Fund Monthly Progress Report*, Sept. 1990. Office of Underground Storage Tanks, U.S. EPA, Washington, D.C.

Pontasch, K.W., Niederlehner, B.R., and Cairns, J., Jr. 1989. Comparisons of single-species microcosm and field responses to a complex effluent. *Environ. Toxicol. Chem.* **8**:521-532.

Porter, K.G., and Feig, Y.S. 1980. the use of DAPI for identifying and counting aquatic microflora. *Limnol. Oceanogr.* **25**(5):943-948.

Pratt, J.R. 1990. Aquatic community response to stress: prediction and detection of adverse effects. In W.G. Landis and W.H. van der Schalie, eds., *Aquatic Toxicology and Risk Assessment: Thirteenth volume. ASTM STP 1096.* American Society for testing and Materials. Philadelphia. pp. 16-26.

Press, W.H., Flannery, B.P., Teukolsky, A.A. and Vetterline W.T. 1990. *Numerical Recipes in C, the art Scientific Computing.* Cambridge University Press, New York, NY.

Rheinheimer, G. 1985. *Aquatic microbiology.* Third edition. John Wiley and Sons, New York.

Schindler, D.W. 1987. Detecting ecosystem responses to anthropogenic stress. *Can. J. Fish. Aquat. Sci.* **44**:5-25.

Shaw, J.F.H., and Prepas, E.E. 1990. Exchange of phosphorous from shallow sediments at nine Alberta lakes. *J. Environ. Qual.* **19**:249-256.

Sloof, W., van Oers, J.A.M., and de Zwart, D. 1986. Margins of uncertainty in ecotoxicological hazard assessment. *Environ. Toxicol. Chem.* **5**:841-852.

Smith, E.P., Pontasch, K.W. and Cairns, J., Jr. 1990. Community similarity and the analysis of multispecies environmental data: a unified statistical approach. *Water Res.* **24**(4):507-514.

Song, H.G., Wang, X., and Bartha, R. 1990. Bioremediation potential of terrestrial fuel spills. *Appli. Environ. Microbiol.* **56**(3): 652-656.

Southerland, E., Kravitz, M., and Wall, T. 1992. Management framework for contaminated sediments (U.S. EPA Sediment Management Strategy). In G.A. Burton, Jr., ed., *Sediment Toxicity Assessment.* Lewis Publishers, Boca Raton, Fl. In Press.

Stay, F.S., Katko, A., Rohm, C.M., Fix, M.A., and Larsen, D.P. 1988. Effects of fluorene on microcosms developed from four natural communities. *Environ. Toxicol. Chem.* **7**:635-644.

Stay, F.S., Katko, A., Rohm, C.M., Fix, M.A., and Larsen, D.P. 1989. The Effects of atrazine on microcosms developed from four natural plankton communities. *Arch. Environ. Contam. Toxicol.* **18**:866-875.

Taub, F.B., and Read, P.L. 1982. Standardized aquatic microcosm protocol: final report contract no. 223-80-2352, vol. II Food and Drug Administration. Washington, D.C.

Touart, L.W. 1988. Aquatic mesocosm tests to support pesticide registrations. Hazard Evaluation Division. EPA 540-09-88-035. U.S. Environmental Protection Agency, Washington, D.C.

Urban, D.J., and Cook, N.J. 1986. Hazard Evaluation Division, Standard Evaluation Procedure, Ecological Risk Assessment. EPA 540/9-85-001. U.S. Environmental Protection Agency, Washington, D.C.

United States Environmental Protection Agency. 1987. An overview of sediment quality in the United States. EPA 905/9-88-002. Office of Water Regulations and Standards, Washington, D.C., and EPA Region 5, Chicago, ILL.

U.S. Coast Guard. 1982-83. Polluting incidents in and around U.S. waters. COMDTINST M16450.2F. Washington, D.C.

Wiederholm, T., and Dave, G. 1989. Toxicity of metal-polluted sediments to *Daphnia magna* and *Tubifex tubifex*. *Hydrobiologia*. **176/177**:411-417.

Zar, J.H. 1984. *Biostatistical Analysis*. Second edition. Prentice Hall, Inc., Englewood Cliffs, NJ.

APPENDIX I

Water Quality Methods

Alkalinity

Alkalinity was measured by potentiometric titration using an Orion model 231 pH meter and an 81-02 electrode calibrated with pH 4 and 7 buffers. Samples were titrated with 0.02 N HCL to pH 4.5 ± 0.05 and the volume of acid recorded. Titration was then continued to pH 4.2 ± 0.02 and recorded. Alkalinity was then calculated as:

$$mgCaCO_3/l = \frac{(2A-B)\times 0.02N\times 50{,}000}{mlsofsample}$$

Where: A= ml required to reach pH 4.5 ± 0.05
B= ml required to reach pH 4.2 ± 0.02
0.02N= Normality of the HCL

Conductivity

Conductivity was measured using a Lamotte DA-LR-699 conductivity meter calibrated to the manufacturers specifications.

Dissolved Oxygen

All dissolved oxygen measurements were obtained using a YSI model 57 dissolved oxygen probe air calibrated to the manufacturers specifications, correcting for both pressure and temperature when possible. This probe was accurate to within ± 0.1 mg/l. Separate probes were used for treated and non treated groups.

Hardness

Hardness was determined by titrating duplicate 50 ml aliquots of samples with 0.01 N EDTA (standardized with 1 mg/l standard calcium solution in the presence of eriochrome Black T indicator (EBT) at pH 10(ammonia buffer)). Titration's were carried out with the pH adjusted to 10 by the addition of an ammonia buffer and ~ 2 grams of EBTindicator. Total hardness was then calculated as:

$$\frac{A\times B\times 1000}{mlsofsample}$$

Where: A = ml of EDTA
B = mg $CaCO_3$ equivalent to 1.00 ml of EDTA
equivalence of EDTA: 1 mg $CaCO_3$

Duplicate values were then averaged and reported.

pH

pH was determined using portable hand-held Piccolo pH meters (HI 1280) calibrated with pH 7 and 10 buffers according to the manufacturers specifications. Readings were obtained using separate meters for treated and non treated groups and were accurate to within ± 0.01 pH units.

Turbidity

Turbidity was measured using a Spectronic 20 calibrated with T82MV.

SUBLETHAL TOXICITY OF COPPER SULFATE TO THE INTERTIDAL SEA ANEMONE, *ANTHOPLEURA ELEGANTISSIMA*

by

Lester W. Keel

SUBLETHAL TOXICITY OF COPPER SULFATE TO THE INTERTIDAL SEA ANEMONE, *ANTHOPLEURA ELEGANTISSIMA*

by

Lester W. Keel

Accepted in Partial Completion

of the Requirements for the Degree

Master of Science

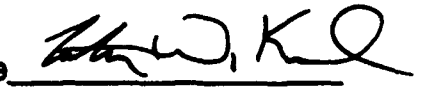
Dean of Graduate School

Advisory Committee

Chair

Master's Thesis

In presenting this thesis in partial fulfilment of the requirement for a master's degree at Western Washington University, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this thesis is allowable only for scholarly purposes. It is understood, however, that any copying or publication of this thesis for commercial purposes, or financial gain, shall not be allowed without my written permission.

Signature 

Date Feb. 5, 1994

Bellingham, Washington 98225 (206) 650-3000

SUBLETHAL TOXICITY OF COPPER SULFATE TO THE INTERTIDAL SEA ANEMONE, *ANTHOPLEURA ELEGANTISSIMA*

A Thesis

Presented to

The Faculty of

Western Washington University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Lester W. Keel

January 1994

ABSTRACT

This study is an initial evaluation of the intertidal sea anemone *Anthopleura elegantissima* as a candidate species for use in marine toxicity monitoring efforts. Many characteristics make *Anthopleura elegantissima* a potentially useful species for *in-situ* or laboratory monitoring of chemical pollution in the intertidal. These include: relatively large size, wide geographic distribution, low mobility, endosymbiotic algae, clonal reproduction, and exposed tissues.

To evaluate this potential, anemones collected from one aggregation were exposed to copper sulfate in filtered sea water for 24 days to determine acutely toxic concentrations. Anemones were acclimated and exposed in the laboratory at 10° C and a 12:12 hour photoperiod. Anemones were fed chopped fresh *Mytilus edulis* tissue, and test media was renewed every three days.

The rangefinding test resulted in a median effective concentration of 1350 µg/L copper. A subsequent 48 day sublethal exposure experiment yielded significant reductions in anemone growth, tentacle extension frequency, and feeding frequency. Endosymbiotic zooxanthellae division was stimulated at 250 µg/L. Copper was bioaccumulated linearly with dose, without apparent regulation. The lowest observed effective concentration for percent weight gain was 175 µg/L copper. These results indicate that *A. elegantissima* is hardy in the laboratory, easily obtainable, and exhibits sublethal effects at concentrations well below that of the copper sulfate EC_{50}. *Anthopleura elegantissima* appears to be a potentially useful biomonitoring species, and further test development is warranted.

ACKNOWLEDGEMENTS

Countless thanks are due to my advisor, Wayne G. Landis, for his patient encouragement, excellent advice, and logistical support. I am indebted to April Markiewicz, and Lisa Holmquist of the Institute of Environmental Toxicology and Chemistry, Western Washington University, for their gracious help. I also wish to thank Gisèle Muller-Parker for her advice, enthusiasm and laboratory accomodations. My appreciation goes to Jack Hardy, and Robin Matthews for ongoing advice and support. I wish to thank Don Williams, and Richard Fonda, of the Biology Department, Western Washington University, for their contributions of advice and encouragement. Douglas Doolittle and Dennis Bohrer, also of the Biology Department, were notably helpful. Others who have greatly helped with this project and to whom I am extremely grateful are Keith Golder, Alan Verde, Mike Roze, and David Tapley.

I am especially grateful to my wife Sheri, and my daughters, Haley, Sydney and Marley, for their inspiration, incredible patience, and loving support during this period.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Biomonitoring involves the use of biological system performance to provide information about pollutant concentrations or impacts on a particular ecosystem. Many marine biomonitoring efforts rely on effects measurements from only one species or utilize bioaccumulation measurements from shelled organisms such as molluscs and crustaceans. This strategy has several shortcomings, notably, the reliance on the responses of a limited number of species, the interference to uptake of pollutants by the shell, and cessation of pumping in molluscs. The use of *Mytilus edulis* in biomonitoring has also been problematic due to irregular reproduction cycles (U.S. EPA, 1989). In addition, the large reliance on bioaccumulation as an ecosystem management endpoint neglects the importance of biological dose response relationships. Thus, there is a need for the additional development of intertidal monitoring species amenable to biological effects measurement and free from interferences that may mitigate accumulation of toxicants.

As an adjunct to existing biomonitoring programs, I have examined the toxicity of copper sulfate to the intertidal sea anemone *Anthopleura elegantissima*. The evaluation consisted of a short-term range finding test measuring the acute toxicity of copper sulfate, and measurements of physiological, behavioral , symbiotic, and bioaccumulation responses after a 48 day sublethal exposure period. Attributes of the species that make it a potentially useful biomonitoring organism include its sessile lifestyle, long lifespan, clonal reproduction, relatively large size, symbiotic associations with unicellular algae, wide distribution, and the abundance of scientific information about the species. Measurement endpoint selection was made on the basis of the mechanistic toxicology of copper, usage of endpoints in other monitoring protocols and in accordance with the biology of *A. elegantissima*. The experimental results are evaluated with respect to the potential utility of *A. elegantissima* as a species useful for marine biomonitoring efforts.

BIOMONITORING

Biomonitoring programs have included both pollutant tissue burdens and biological effects measurement. A biomonitoring method may examine pollutant impacts at several levels of biological utilization and organization: chemical uptake, transformation and degradation, site of action effects,

biochemical responses, physiological and behavioral responses, and population, community and ecosystem effects. The efficacy of a biomonitoring program depends on both the minimum concentration of toxicant that may be reliably detected (sensitivity), and the amount of delay between the appearance of putatively toxic conditions, and the presentation of the endpoint response measured (Van der Schalie, 1986).

Regulatory Mandates for Biomonitoring

Biomonitoring may be used retroactively to assess ecosystem damage after a pollution event, for example in natural resource damage assessments prescribed by the Oil Pollution Act of 1990 (U.S.C. 33§2761), and the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (U.S.C. 42§9601 *et seq.*). Alternatively, toxicity testing and pollutant uptake tests may be used to predict pollution effects *a priori* such as during chemical screening mandated by the Toxic Substances Control Act of 1976 (U.S.C. 15§2603). Biomonitoring is also used to routinely assess effluent toxicity under the Federal Water Pollution Control Act of 1987 (U.S.C. 33§1254). A fourth purpose for Biomonitoring is also associated with programs that survey ecosystems to determine long-term trends in ecosystem health, for example the use of sentinel organisms in the field or in the laboratory to detect large scale spatial and temporal trends in toxicity or bioaccumulation.

There are several national ecosystem monitoring programs administered by the U.S. Environmental Protection Agency (USEPA) that are pertinent to this research: the Environmental Monitoring and Assessment Program, and the Status and Trends program. In addition, the Puget Sound Water Quality Authority administers biological surveys in Puget Sound such as the Puget Sound Estuary Program, and the Puget Sound Ambient Monitoring Program.

The regulatory use of biomonitoring for contaminants in aquatic environments officially began in 1984 when the Office of Water of the EPA issued a 'Policy for the Development of Water Quality Based Permit Limitations for Toxic Pollutants' (Federal Register, 1984). This policy established a commitment by the regulatory community to use an integrated strategy consisting of both biological and chemical methods to address toxic and conventional pollutants. The 1987 amendments to the Federal Water Pollution Control Act significantly strengthened this position. While this program is meant

to assess "end of pipe" toxicity in conjunction with water quality criteria, the USEPA regulates new and existing chemicals, and genetically engineered organisms through the Toxic Substances Control Act (TSCA). The Toxic Substances Control Act may regulate the chemical if there is an unreasonable risk to human health or the environment. Under TSCA, toxicity testing information is collected in a tiered fashion as part of a hazard assessment that addresses the life cycle of the substance and the risks to aquatic and terrestrial biota.

Biomonitoring Program Design

Well-designed biomonitoring programs make use of a variety of toxicity tests in conjunction with physicochemical data to provide an assessment of the overall impact of a toxicant on an ecosystem. Toxicity tests are used as components of a conceptual model of environmental assessment that includes toxicant loading and fate, body burdens, organismal and population effects, ecosystem changes, and the socially derived ecosystem functions and values that are to be protected (White, 1984).

Several tradeoffs and problems must be addressed while designing a biomonitoring program. A shortcoming of many previous biomonitoring surveys has been collection of data solely on tissue toxicant burden. While this is an indication of exposure, toxicity must be estimated from the data. Measurements of biological responses to toxicants are an advantage over physicochemical monitoring, as chemical concentration data alone are not sufficient to predict toxicity. Biological sensitivities may exist at concentrations below analytical detection limits, and toxicity may be altered by ecosystem factors, interactions of chemicals, and site-specific water quality.

A tradeoff exists between the reliability of a biological effects measurement for detecting a toxic effect, and detecting a specific toxicant, or between a measurement that gives an early indication of toxicity, and a response that may become manifest later but with fewer "false positive" responses. Extrapolations must often be made between taxa, measurement endpoints, laboratory responses and field responses, levels of biological organization, and geographic locations or seasons of the year. Finally, there is the possibility of acclimation to the toxicant attenuating the biological effects dose response or bioaccumulation.

Selection of Aquatic Biomonitoring Species

Species selection is only one part of the overall design of a biomonitoring protocol but it relates to many of the design problems listed above. It is impossible for one species to provide enough information to make conclusions regarding the greater ecosystem (Cairns, 1986). It is, therefore, preferable if biomonitoring schemes make use of a variety of species at several trophic levels. The following is a list of criteria for selection of appropriate sentinel organisms to be used in aquatic field surveys, although many of the criteria also apply to the development of laboratory toxicity tests (Phillips, 1980; Van der Schalie, 1986).

1. The species must integrate pollutant effects over time.

2. The species should be sedentary or sessile to be representative of a particular geographic area.

3. The species should be common and abundant for ease of collection.

4. The species should be large enough to provide sufficient tissue for analysis.

5. The age or size should be sufficient to allow sampling of more than one year class.

6. The species should be tolerant of laboratory conditions.

7. The species should be tolerant of lower salinity and higher temperatures (estuarine adaptation).

8. There should be a correlation between water concentration and the organism body burden.

We may add to this list the requirement that the species tested be endemic to the locations receiving the subject toxicant for at least part of its life cycle. Some states require the use of indigenous species while the EPA allows departure from species specified by regulation only under special

circumstances (Webber, 1991). In general, the suitability of a species for evaluating pollution depends on its ability to reliably reflect the ecosystem from which it is sampled. Alterations in environmental pollutant levels should result in reproducible, quantifiable, and meaningful changes in physiological, biochemical, or morphological characteristics in the species under examination.

Marine Biomonitoring Species

Few marine species are used in regulatory toxicity testing programs, and most species tested are fishes or arthropods. The most commonly used are the sheepshead minnow (*Cyprinodon variegatus*), several silverside species (*Menidia sp.*), the bay mysid (*Mysidopsis bahia*), and the copepod (*Acartia tonsa*) for acute tests. The bay mysid and the sheepshead minnow are the preferred species for lifecycle tests (Hansen, 1984). Echinoderm gamete, embryo, and adult tests are in common use (reviewed by Bay et al., 1993). Tests utilizing reproduction of macroalgae, and growth of microalgae have proven successful (Thursby et al., 1993). Bowmer et al. (1986) developed toxicity tests utilizing a variety of responses in the brittle star *Amphiura filiformis.*

The most widely used *in-situ* biomonitoring species is the common mussel *Mytilus edulis,* which has been employed both in the United States and in other parts of the world in the National, International, and California Mussel Watch programs (Goldberg et al., 1978; Martin and Severeid, 1984; U.S. EPA, 1989). Typically, *Mytilus edulis* suspended in cages within the water column are periodically sampled for survival, shell growth, scope for growth, and bioaccumulation determinations (U.S. EPA, 1989).

TOXICITY TESTING WITH CNIDARIANS

Few cnidarians have been used to study the effects of environmental toxicants, although hydroids have received some attention (reviewed by Stebbing and Brown, 1984). Specific growth rate, morphological changes, behavior, and lysosomal hydrolase activity have been applied as measurement endpoints in hydroids exposed to metals and other toxicants (Karbe, 1972; Stebbing, 1976; Moore and Stebbing, 1976; Moore, 1980; Houvenaghel, 1984). Stebbing (1976) developed an assay for metal toxicity using the inhibition of colonial growth rate in the marine colonial hydroid *Campanularia flexuosa* (*Laomedea flexuosa*). Stebbing observed inhibition of colonial growth rate after eleven days exposure to 10-13 µg/L copper. Karbe exposed the marine

colonial hydroid *Eirene viridula* to Cu, Pb, Zn, Hg, and Cd over short (2-3 weeks) and long (3 months) periods (1972). Tissue disintegration was evident within a few hours at 3000 µg/L copper, and morphological changes occurred at 60 µg/L copper. The threshold concentration for acute effects was from 30 to 60 µg/L copper in Karbe's experiment. Houvenaghel (1984) examined inhibition of feeding rate in the hydroid *Hydractinia echinata* exposed to a variety of pollutants. The freshwater cnidarian *Hydra attenuata* was used as a first tier evaluation of teratology (Johnson, 1983).

Campanularia flexuosa exposed to field samples from a polluted estuary showed morphological changes and increases in gonozooids which were positively correlated with dissolved copper and cadmium concentrations (Stebbing et al., 1983).

In hydroids, starvation, low temperatures, and presumably other environmental stressors cause tissue degeneration and reduced growth (Moore and Stebbing, 1976). Degeneration is reversed when environmental conditions improve. Lysosomal hydrolase activity from *C. flexuosa* exposed to copper was found to be a more sensitive endpoint than tissue degeneration in response to copper intoxication (Moore and Stebbing, 1976). Lysosomal hydrolase plays a role in tissue degeneration.

The anthozoan *Anemonia viridis* regulated copper uptake over 5 day exposures to 50 and 200 µg/L copper (Harland and Nganro,1990). Harland et al. (1990) measured significant zinc and cadmium accumulations during laboratory exposures of the sea anemones *Anemonia viridis* and *Actinia equina*. Brown and Howard (1985) measured Cu uptake during 7 day laboratory exposures of *A. equina* and *A. viridis*. Both species exhibited uptake only at the highest concentration tested (200 µg/L).

Anthopleura elegantissima as a Biomonitoring Species

With respect to the species selection criteria outlined above, several attributes of *Anthopleura elegantissima* make it a potentially useful biological indicator of chemical stress in the intertidal. It is distributed from Alaska to Southern California between 0 and 4.5 feet above mean lower low water level (Hand, 1955). This wide distribution facilitates long-range site comparisons. Being intermediate in mobility between motile predators and sessile suspension feeders, the species remains localized in the intertidal with movements related to secondary habitat selection (Sebens, 1982). Low motility

allows field data to be representative of geographically discrete areas. Moreover, the absence of a shell creates the potential for continuous dermal exposures of *A. elegantissima* to toxicants borne by aerial deposition, water column , and surface microlayer pathways.

The most common growth form in the intertidal is the clonal aggregation. A solitary growth form exists lower in the intertidal, often subtidally (Francis, 1979). Sexual reproduction through the release of gametes occurs approximately annually (Sebens, 1981a). Clonal aggregations are formed asexually by longitudinal fission of individuals; making the aggregation genetically homogeneous. Thus, repeated collection of genetically identical individuals is possible. Experimentation with monoclonal replicates is useful for assessing environmental factor interactions with phenotypic variation. That is, the responses of genetically identical organisms may be used to distinguish between genetic and environmental factors affecting toxicity. Moreover, experimental variance is generally lessened. One caveat to the use of monoclonal replicates in toxicity testing is that the response of one genotype is not necessarily reflective of the average species response.

Anthopleura elegantissima exposed to sunlight typically have endosymbiotic algae in their gastrodermal cells (Hand, 1955). Endosymbionts may be an unknown species of *Chlorella* , or the dinoflagellate *Symbiodinium californianum*. This symbiotic association creates a venue for assessing toxicity at two trophic levels simultaneously.

Anthopleura elegantissima is a conspicuous species in the coastal rocky intertidal system. In the California rocky intertidal zone, its percent cover (3%) is second to the barnacle *Chthamalus fissus* (4%), and ranks second in biomass (24 g dry wt $\cdot m^{-2}$) to the mussel *Mytilus californianus* (49 g dry wt$\cdot m^{-2}$) (Littler, 1980). Consistent with the species selection criteria above, *A. elegantissima* is an important component of the intertidal ecosystem; Fitt et al. (1982) calculated gross primary productivity for zooxanthellate *A. elegantissima* and found it similar to that of intertidal seaweed populations (48-151 g $C\cdot m^{-2}\cdot yr^{-1}$). *A. elegantissima* is preyed upon by the aeolid nudibranch, *Aeolidia papillosa*, (Macfarland and Muller-Parker, 1993), and the sea star *Dermasterias imbricata* (Sebens, 1983), and possibly various fish, bird and mammal species (Ates, 1991).

Development of toxicity testing methodologies utilizing *Anthopleura elegantissima* benefit from the abundance of prior research with this species.

The reproductive biology is well defined (Ford, 1964; Sebens, 1980; 1981a,b; 1982). Feeding strategies, energetics, biochemical composition, and the symbiotic interactions have been explored (Sebens, 1981b; Jennison, 1979; Zamer and Shick, 1987, 1989; Shick and Dykens, 1984;). Aspects of oxyradical metabolism have been studied (Dykens et al., 1992; Dykens and Shick, 1982). This research also attests to the hardiness of the species in the laboratory.

TOXICOLOGY OF COPPER

Copper chemistry in marine waters is influenced by both biotic accumulation causing surface enrichment, and physical processes such as upwelling and ligand formation causing desorption from sediments (Boyle and Edmond, 1975). Boyle reviewed the available data on copper in seawater and found that concentrations ranged from 0.15 µg/L in the open sea, to about 1.0 µg/L in polluted estuaries (1979). Copper was selected for this research not only for its relevance to coastal pollution, but because it is routinely utilized as a reference toxicant for toxicity test development and laboratory quality assurance (Jop, et al., 1993).

Copper is an essential element, utilized as a prosthetic group in various enzymes such as superoxide dismutase, ceruloplasmin, tyrosinase, laccase, cytochrome *c* oxidase, ascorbate oxidase, and lysyl oxidase. Copper is also in the oxygen-carrying molecules hemocyanin and plastocyanin. Copper plays a role in essential functions such as electron transport, collagen synthesis, melanin formation, hemoglobin synthesis, and amino acid metabolism (Ettinger, 1984). Many copper-containing enzymes function as oxidases, reducing molecular oxygen to water while oxidizing a variety of inorganic and organic substrates.

As a result of the hormetic nature of copper, organisms have developed regulatory pathways that control cellular copper metabolism. In general, these pathways employ a series of metal binding ligands. In mammals, as copper enters the cell it binds first to reduced glutathione, then is transferred to metallothioneins where it is stored or subsequently transferred to copper enzymes (Freedman et al., 1989). Albumin and histidine in plasma also play roles in copper homeostasis (Ettinger, 1984). Animals exposed to copper typically have greater levels of thionein-like copper binding proteins (Viarengo et al., 1981). Copper tends to concentrate intracellularly in lysosomes

(Sternlieb and Goldfischer, 1976). Lysosomes are formed by vesiculation from Golgi bodies and are routinely used in degradative metabolism (catabolism). High levels of heavy metals, including copper, can result in a loss of lysosomal membrane stability and the accumulation of lipofuscin granules within the lysosomes (Regoli, 1992). Before excess copper can cause toxicity it must enter the cell at rates beyond the range of control afforded by the various pathways listed above. Alternatively, endogenous copper may escape the normal metabolic pathway, as in inborn errors of metabolism such as Wilson's disease (Ettinger, 1984).

Copper may cause toxicity by a variety of mechanisms. First, exogenous ligands may enter the cell and extract copper from normal copper binding sites thereby inactivating copper enzymes and/or creating new cytotoxic copper-ligand species. Second, copper may displace beneficial metals from the active sites of enzymes. Third, copper may bind to a deactivating (or activating) site on enzymes or nucleotides, especially at nucleophilic groups. Copper commonly forms complexes with organic molecules at groups containing sulfur, oxygen, and nitrogen, it has a particularly high affinity for sulfhydryl groups (Martin, 1986). Fourth, as a transition metal active in redox reactions with both Cu^{+1} and Cu^{+2} redox states available intracellularly, copper-ligand complexes may participate in reactions that spawn damaging free radicals (Petering and Antholine, 1988). Finally, copper binding and redox cycling in conjunction with glutathione can deplete cellular glutathione content thereby upsetting the cellular redox status (Viarengo et al., 1990). These latter two aspects are taken up more fully in the following sections.

Aside from steric modifications imposed by nonspecific binding of copper to biomolecules such as proteins, and nucleotides, some ligand-copper complexes effect damage through free radical production. As the main damaging species produced is the extremely reactive hydroxyl radical, much of the free radical damage is located at or very close to the copper complex. A free radical is any species that has one or more unpaired electrons, such as the hydrogen atom, diatomic oxygen, and most transition metals, including copper. Less reactive secondary radicals may travel greater distances to inflict damage elsewhere. Lipid radicals can arise and participate in a peroxidative chain reaction that damages membrane lipids and membrane proteins (Chan et al., 1982; Borg and Schaich, 1984; Bus and Gibson, 1979). Copper-imposed free radical toxicity is mediated primarily by activated oxygen species.

Pertinent to this research are the activated oxygen species, as copper has been shown to catalyze the formation of destructive oxyradicals (Samuni et al., 1981; Aruoma et al., 1991). Activated oxygen species consist of the one, two, and three electron reduction products of O_2, namely the superoxide anion radical (O_2^{-}), peroxide anion (O_2^{2-}), and the hydroxyl radical (·OH). The hydroxyl radical reacts extremely rapidly with most biomolecules: sugars, amino acids, lipids, DNA, phospholipids, and organic acids through three different mechanisms: hydrogen atom abstraction, addition, and electron transfer.

Copper undergoes redox cycling with hydrogen peroxide and superoxide anion to catalyze the formation of hydroxyl radicals according to the metal catalyzed Haber-Weiss reaction as follows (Samuni et al., 1981):

$$2O_2^{-}\cdot + 2H^{+}\cdot \quad <----> \quad H_2O_2 + O_2 \qquad (a)$$

$$H_2O_2 + [Cu^{(1+)}] \quad <----> \quad \cdot OH + OH^{-} + [Cu^{(2+)}] \qquad (b)$$

$$O_2^{-}\cdot + [Cu^{(2+)}] \quad <----> \quad [Cu^{(1+)}] + O_2 \qquad (c)$$

$$O_2^{-}\cdot + H_2O_2 \quad <----> \quad O_2 + \cdot OH + OH^{-} \qquad (d)$$

Superoxide anion plays a dual role of reducing copper (equation c), and dismutating spontaneously or with superoxide dismutase (SOD) to form hydrogen peroxide (equation a). The superoxide anion mediated redox cycling of copper shown in equation (c) can be replaced by other reducing agents such as ascorbate which also produces hydrogen peroxide during its oxidation (Shinar et al., 1983). Reaction (b) is the Fenton reaction known to occur with both copper and iron salts in *in-vitro* aqueous solution. Reaction (d) shows the net Haber-Weiss equation in the presence of the complexed copper salt catalyst.

Oxygen radicals are produced in aerobic organisms by hemoglobin oxygen transfers, mitochondrial, and photosynthetic electron transport chains (ETC), and through organic xenobiotic redox cycling via NAD(P)H-dependent reductases (mixed function oxidase system) (Foote, 1976; Jewel and Winston, 1989; Livingstone et al., 1990). Hydrogen peroxide is formed by the spontaneous dismutation of superoxide anion, or catalytically with the enzyme

superoxide dismutase. Superoxide formed during oxidation of $Cu(^{+1})$glutathione, and during the oxidation of reduced glutathione may also dismutate to hydrogen peroxide (Minkel et al., 1980). Thiols, in general, oxidize rapidly in the presence of copper and other transition metals with a resultant production of hydrogen peroxide (Nath and Salahudeen, 1993). Finally, free radicals may arise through photooxidations of heterocyclic compounds such as histidine (Foote, 1980).

In summary, the coupling of redox cycles with an electron donor via an autoxidizing substrate yield the dismutation product hydrogen peroxide. Hydrogen peroxide is converted to the hydroxyl radical through a metal catalyzed heterolytic cleavage reaction. The metal is subsequently reduced by superoxide anion or another reducing agent (e.g.. ascorbate or glutathione). Hydroxyl radicals react within a few angstroms of the metal to inactivate biomolecules through a free radical chain reaction. Secondary radicals such as the metastable $HO_2\cdot/O_2^-$, and carbonyl radical species can pass through membranes to cause damage at a distance as can lipid radical formed into hydroperoxides. Hydrogen peroxide may also diffuse to other compartments to react at other metal centers. This sequence can generally be mitigated through removal of activated oxygen species by superoxide dismutase or catalase.

The theoretical evidence for metal-catalyzed free radical toxicity is supported by measurements of antioxidant induction, and lipid peroxidation and other by-products of radical chain reactions in copper-exposed biological systems. Copper-induced damage to calf thymus DNA was characteristic of hydroxyl radical caused lesions in an *in-vitro* system containing copper, hydrogen peroxide, and ascorbic acid (Arouma et al., 1991). Inhibition of acetylcholine esterase activity *in--vitro* showed a similar dependence on copper ions in conjunction with ascorbate and hydrogen peroxide (Shinar et al., 1983). *Mytilus galloprovincialis* exposed to copper showed increased levels of malondialdehyde and other aldehydic compounds. These compounds are by-products of lipid radical chain reactions. An increase in lysosomal lipofuscin granules was noted in the digestive gland as a possible detoxification modality. Lipofuscin is a by-product of lipid peroxidation. The free radical detoxification compound glutathione was significantly depressed in exposed mussels (Viarengo et al., 1990).

There is an array of enzymes and non-enzymatic defenses to free radical toxicity. For example, catalase (EC 1.11.1.6), superoxide dismutase (EC

1.15.1.1), ascorbate peroxidase and glutathione peroxidase (EC 1.11.1.7), ascorbic acid, glutathione, and the fat soluble vitamins alpha-tocopherol and beta-carotene all attenuate the formation of oxyradical species (Winston, 1991).

ENDPOINT SELECTION

Appropriate biological effects, or endpoints, must be selected for measurement in conjunction with the species selection. Similar to species selection, criteria exist for endpoint selection that generally address variability, sensitivity, ease of measurement, and integration of effects. The majority of field assessment programs measure toxicant body burdens rather than biological effects. Martin and Severeid (1984) listed the criteria below for biomonitoring endpoints used as part of the California State Mussel Watch Program.

1. The biological response indicator should be quantitatively influenced by toxic pollutants or other environmental stressors.

2. The biological response indicator should compensate for natural environmental stressors and thus respond only to stress induced by toxic pollutants.

3. The biological response indicator should have a significant biological or ecological meaning (survival, growth, recruitment, reproduction).

4. The biological response indicator should be a quantitative statement of sublethal or chronic impacts of pollution.

5. The biological response indicator should be reasonably easily measured in the field or laboratory.

6. If an adverse effect is measured at the organismal or population level, the biological response indicator should be interpretable at other levels of organization: subcellular, cellular, and ecosystem.

7. The biological response indicator should be referable to historical biological and chemical information and data sets.

SELECTED MEASUREMENT ENDPOINTS

The endpoints selected for measurement in this research were chosen on the basis of the characteristics of *A. elegantissima*, the mode of action of copper, and the endpoint selection criteria outlined above. Anemone weight and behavior were measured over the course of the exposure period. Catalase activity, endosymbiotic zooxanthellae density and the percent in division, photosynthesis and respiration, and copper bioaccumulation were measured at the conclusion of the exposure period.

Growth

Growth measurements integrate the processes of energy acquisition and expenditure, and mechanistic toxicology. Copper intoxication can cause energetic diversion from growth to cell repair or detoxification processes. Other research has demonstrated inhibition of growth with metal exposure. Coho salmon (*Oncorynchus kisutch*) exposed to sublethal levels of copper exhibited initially decreased rates of growth, but later became acclimated (Buckley, et al 1982). The growth inhibition was accompanied by reduced feeding rate. Hydranth growth rate in the freshwater hydroid *Hydra littoralis* was inhibited by copper (Karbe 1984). Brown and Howard (1985) found reduced growth rate among branching corals exposed in the field to tin smelter effluent. They did not find this effect in the massive corals, such as *Porites*. Growth rate has been a common endpoint in marine hydroid toxicity tests. Karbe (1972) found growth inhibition in *Eirine viridula* exposed to copper, mercury, zinc, and cadmium. Colonial growth rate of the marine hydroid *Campanularia flexuosa* was inhibited by 10 µg/l copper (Stebbing 1976; Stebbing 1979). Below this level copper exposure resulted in increased gonozooid (reproductive polyp) production over gastrozooid (feeding polyp) production. Initially decreased specific growth rates, but with recovery, were noted on exposure of *C. flexuosa* to reduced salinity (Stebbing 1981). A similar pattern of gonozooid production was also noted.

In the case of *A. elegantissima*, which has indeterminate growth and lifespan, attainment of larger body size yields greater sexual and asexual reproductive output. As gonad volume, and presumably gamete production, is directly related to body size, larger size will result in greater sexual reproductive output (Sebens, 1981a). *Anthopleura elegantissima* reproduces asexually by

longitudinal fission (Hand, 1955). The timing of fission appears to be related to prey capture area and energetic requirements with larger anemones dividing more frequently than smaller ones (Sebens, 1979; Sebens, 1982). Only *A. elegantissima* greater than 1.0-1.2 cm basal diameter, approximately 2 years after settlement, will divide. In the rocky intertidal, asexual reproduction can be advantageous in terms of secondary competition for space and for the amplification of successful genotypes (Paine 1966; Shick et al., 1979). Larger body size can also reduce the effects of predation (Francis, 1979). Thus, inhibition of growth has important ecological consequences.

Behavior

Behavioral changes that involve feeding, reproduction, predator/ toxicity avoidance, or learning can adversely affect survival and fitness. Behavioral toxicity can result from impaired sensory function or nerve signal transmission. *Anthopleura elegantissima* possess gastrodermal and epidermal nonpolar nerve nets that are linked across the mesoglea. Coordinated activity is facilitated by differently organized neuronal conducting systems linking sensory cells, muscles, and neural concentrations (pacemakers) (Shick, 1991). Sensory capabilities are carried out by scattered receptor cells, typically with modified apical flagella. Discharge of the nematocysts is stimulated by mechanical contact of the food item, or by chemical stimuli at ciliary cones at the apex of the nematocyst capsule (Mariscal, 1974). Nematocyst discharge also appears to be mediated by nervous system control as nematocyst discharge is reduced after feeding (Shelton, 1982).

The behavioral repertoire of sea anemones is limited compared to mammals and fish, however, there are simple and easily observable behaviors related to survival. Anemones commonly retract the gastric column when mechanically disturbed and nematocyst discharge has been shown to increase with continued mechanical stimulation (Conklin and Mariscal, 1976). Patterns of behavior have been established for tentacle and gastric column retraction responses to high light intensities (Shick and Dykens, 1984). Feeding rate in hydroids has been explored as a toxicological endpoint (Houvenaghel, 1984). Fredericks (1976) observed attraction/avoidance responses to an oxygen gradient in *A. elegantissima*. Intraspecific contact between non-clonemates elicits the defensive acrorrhagial response (deposition of nematocyst bearing tissue from specialized areas on the column) (Francis, 1976). This relative

simplicity of behavioral options results from the limited behavioral demands of a passive feeding, sedentary lifestyle, and, in part, from the two dimensional aspect of the nervous system, and the small variety of cell types (Shick, 1991).

Sea anemones extend their tentacles for passive capture of prey or to increase light capture for photosynthesis. Shick and Dykens (1984) showed that the degree of tentacle expansion was inversely related to light intensities above 200-300 $\mu E \cdot m^{-2} \cdot s^{-1}$. Tentacles are frequently retracted during high light intensities with the effect of reducing photosynthesis rate. They suggested that there was no tidal or circadian rhythm to expansion as dim illumination rendered continuous expansion.

Anthopleura elegantissima is an opportunistic feeder and will ingest a large variety of prey, the bulk of which is plant material, molluscs, and crustaceans (Sebens, 1981b). The feeding response is a well defined behavioral pattern with three phases; 1) prey capture by nematocysts and movement of the prey to the mouth, 2) opening of the mouth, and 3) ingestion of the prey. Reduced glutathione, proline and asparagine leaked from prey can elicit preparatory behaviors to the feeding response . Behavioral toxins can interfere with the feeding reaction at a variety of stages (Lindstedt, 1971):

1. tentacle orientation;
2. initiation of the feeding response;
3. continuation of the feeding response;
4. termination of the feeding response.

Copper may exert damage to the nervous system by damaging receptors, or by interfering with synaptic or neuronal transmission. Magnesium ions inhibit synaptic transmission and are often used with invertebrates as an anesthetic. Houvenaghel (1984) measured feeding response of the marine hydroid *Hydractinia echinata* challenged by various toxicants. With polar and ionic substances there was strong and long lasting contraction of the gastric column and tentacles and subsequent inhibition of the feeding response. Feeding activity of the barnacles *Semibalanus balanoides* and *Balanus crenatus* were diminished by exposure to 80 ppb Cu (Powell and White, 1989). Copper in the range of 20.8-25.6 µg/l caused complete cessation of *Mytilus edulis* pumping rate (Redpath and Davenport, 1988). The effect of copper on *Mytilus* was an "all or none" response with pumping rate stoppage caused by

shell adduction. Chromium at 1.0 mg/l significantly inhibited filtration rate in the bivalves *Mytilus edulis* and *Mya arenaria* (Capuzzo and Sasner, 1977). Chemoreceptors of the blue crab *Callinectes sapidus* were significantly damaged by 100 µg/l copper (Bodammer, 1979).

Photosynthesis and Respiration

Photosynthesis and respiration rates are integrated measurements of energy flow and each may be diminished or enhanced according to the dose and mechanism of toxicity. They are integrated in the sense that any number of toxicant induced lesions or forced metabolic pathway biases could result in stimulation or depression of these two processes. Copper, and even compounds that are not nutritionally important, may cause stimulation of respiration in hydroids at low concentrations (Stebbing, 1976).

In general, photosynthesis and respiration rates are extremely variable over the course of a day or year depending on body size, habitat, behavior, temperature, nutritional condition, and season. Excursions of photosynthesis and respiration rates that are toxic must, by definition, diminish survival or reproduction. Changes in photosynthesis and respiration would most likely be useful as "alarm parameters", giving early warning of toxicity (Zachariassen et al., 1991).

Photosynthetic light saturation for *A. elegantissima* occurs between 125-350 $\mu E \cdot m^{-2} \cdot s^{-1}$ (Fitt et al., 1982). Light saturation is the point at which increasing light intensities no longer stimulate photosynthesis. Inhibition of photosynthesis occurs in most phototrophs at high light intensities, but this effect was not found in *A. elegantissima* at intensities up to 1550 $\mu E \cdot m^{-2} \cdot s^{-1}$. Gross photosynthetic rates are correlated with chlorophyll *a* content and zooxanthellae number, and weight-specific gross photosynthesis rates are inversely related to anemone size (Fitt et al., 1982). Notably, Fitt's group did not find differences in maximum photosynthetic rate between fed, starved, or freshly collected *A. elegantissima* (Fitt et al., 1982). High shore and low shore anemones did not differ in chlorophyll *a* content or biomass ratios of zooxanthellae to host (Shick and Dykens, 1984). *Anthopleura elegantissima* varied photosynthetic rates over short time periods through shading of zooxanthellae by contracting or by attaching debris to the column verrucae in response to high light intensity induced hyperoxia and ultraviolet radiation (Shick and Dykens, 1984). This behavior was more frequent among high shore

anemones. Expulsion of endosymbionts may be a long-term adaptation to high light intensity (Pearse, 1974).

Respiration rates in *A. elegantissima* depend on anemone weight, nutritional state, behavior, and emersion periodicity. Oxygen consumption rates decrease with increasing anemone weight and oxygen consumption is highly dependent on feeding state, as fed anemones have nearly twice the rate of starved or newly collected anemones (Fitt et al., 1982). Respiration rates did not change over the course of the day. Anaerobic metabolism is not often utilized as Shick and Dykens (1984) showed intertidal *A. elegantissima* to remain fully aerobic during up to 15 hours emersion in the dark. In high shore anemones, comparable to shading behaviors that minimize photosynthesis, oxygen consumption may be reduced by quiescence. Anemones acclimated to low shore heights did not reduce their activity upon emersion.

Fed *A. elegantissima* received approximately 13% of the carbon fixed by endosymbiotic zooxanthellae (Fitt et al., 1982). This quantity greatly increased for starved or newly collected *A. elegantissima* (45%). Nutritional state also affected gross photosynthesis to respiration ratios (P:R). Starved anemones had P:R in the range of 2.0-3.0 while P:R was usually less than 1.0 in fed anemones. Zamer and Shick (1987) calculated energy budgets for *A. elegantissima* and estimated the translocation of carbon fixed by endosymbiotic zooxanthellae at 41 and 79% for high and low shore anemones respectively. Scope for growth was also different depending on shore height with a higher value due to increased prey capture rates, greater prey absorption efficiencies and reduced metabolic demands due to greater aerial exposure in high shore anemones. Interestingly, a large part of the energy budget was attributed to mucus production (ca. 30%).

Reductions in photosynthetic rate due to copper toxicity has been demonstrated by Rijstenbil and Wijnholds for the marine diatom *Dictylum brightwelli* (1991). Anderson and Morel (1978) also recorded photosynthetic inhibition in the marine dinoflagellate *Gonyaulax tamarensis* exposed to copper. Scott and Major (1972) measured significant heart rate and respiration depression in *Mytilus edulis* at exposures greater than 0.2 ppm (200 µg/L) copper.

Bioaccumulation

Bioaccumulation is the process by which chemicals enter an organism through the diet or by direct absorption through epithelia and respiratory surfaces. Bioaccumulation is determined by chemical analysis of residues or metabolites of the toxicant present in tissues. Bioaccumulation tests have been popular as an indicator of exposure and the availability of toxicants. Compounds that bioaccumulate often tend to concentrate in ascendancy through trophic levels (biomagnification). For bioaccumulation data to yield toxicological usefulness, the data must be equated with biological effects (Peddicord, 1984; Widdows and Donkin, 1991). Bioavailability is dependent on both the target organism, site specific water quality such as pH, salinity, and dissolved organic content. Body burdens are not simple reflections of ambient toxicant concentrations. The regulation of toxicant uptake or induction of detoxification processes makes interpretation of tissue toxicant measurements difficult and hampers the correlation of tissue burdens with toxic effects.

Ionic copper may exist at the +1, +2, and +3 oxidation states, with copper(II) being the most stable in aqueous solution. Copper(III) complexes are relatively rare, and are unstable in aqueous solution. The relative stability of Cu(I) and Cu(II) species in solution, and their biological availability depend on the nature of the ligands present and the composition of the solution. Copper toxicity and uptake are most closely related to availability rather than total copper concentration. Copper availability is dependent on pH, and the presence of organic or inorganic ligands. Sunda and Guillard (1976) demonstrated this dependence by altering copper availability independently of total copper concentration with the use of varying chelator concentrations and measuring toxicity to the marine diatom *Thalassiosira pseudonana* . Lower pH increases toxicity presumably due to an increase in free cupric ion.

Toxicant uptake at the organism level is mediated by anatomy, behavior, size, seasonal cycles, and by the presence of endogenous sequestration pathways, such as metal binding proteins. As uptake is largely a function of surface area, small organisms and those without shells would absorb (and adsorb) a greater mass of toxicant per body weight than organisms partly covered by shells or that are larger with a smaller surface to body weight ratio. This relationship may be overcome by tissues with greater absorptive area, such as gills. Behavior is important in that it includes avoidance of toxic conditions, for instance, cessation of pumping in mussels or escape by

swimming. Also, different feeding methods may place the organism close to toxicant laden sediments or at the upper trophic levels where secondary poisoning is exacerbated by biomagnification. Seasonal cycles of gonad production yield fatty tissues that absorb lipophilic compounds. Membrane transport proteins facilitate uptake of essential nutrients and toxicants may compete for this process or damage their function.

Harland and Nganro (1990) measured copper uptake by the symbiotic sea anemone *Anemonia viridis* . Their study indicated that *A. viridis* regulated uptake over 5 days exposure to 0.05 and 0.2 mg/l copper. Proposed mechanisms for the regulation of copper were the expulsion of zooxanthellae and the production of mucus. Harland et al. (1990) noted that zinc uptake in the anemones *Anemonia viridis* and *Actinia equina* was slight at polluted areas, but that uptake was greatly enhanced in laboratory exposures. This disparity may be due to the presence of chelating compounds in natural waters.

Zooxanthellae Density and Reproduction

The *A. elegantissima* used in this study hosted the endosymbiotic dinoflagellate *Symbiodinium californianum*. Zooxanthellae reside intracellularly within vacuoles of gastrodermal cells. Endosymbiotic zooxanthellae density and number of algae in the process of division (mitotic index) are endpoints that are easily measured from anemones in the field by simple enumeration from preparations of tentacle tissue. These parameters are important, since reductions in zooxanthellae standing crop may reduce the productivity of the symbiosis.

Zooxanthellae density is a function of algal division rate, algal death rate, and expulsion rate. Wilkerson et al. (1983) measured the ratio of dividing zooxanthellae to zooxanthellae density (mitotic index) in *A. elegantissima* freshly collected from Lopez Island, Washington. Wilkerson's results showed asynchronous zooxanthellae division and a mean mitotic index of 2.88%. The approximate doubling time calculated from the mitotic data for the endosymbiotic zooxanthellae population was 11.2 days.

Cnidarians exposed to temperature stress increase the expulsion of endosymbionts (Glynn, 1984). Brand et al. (1986) found the *Gymniodinium* sp. reproductive rate dose response to be stimulated at low concentrations and depressed with higher copper concentrations and reproductive rate was more sensitive to copper than reductions in photosynthetic rates. Elevated heavy

metal concentrations increased expulsion of zooxanthellae in laboratory exposures (Harland and Nganro, 1990). Accumulation of copper in zooxanthellae and subsequent expulsion may be a modality for regulation of copper (Harland and Nganro, 1990).

Catalase Activity

Biochemical endpoints of toxicity (biomarkers), such as enzyme induction/inhibition, stress proteins, and immunoiogical suppression, are typically measurements selected on the basis of knowledge about the mode of action of the toxicant to obtain early indications of damage at the cellular level. Biomarkers tend to give information that is more detailed and specific to the chemical and the organism. Hence, there can be difficulty interpreting biomarker response with regard to organism health or population effects. Biomarkers may be distinguished as either adaptive responses, such as the case of enzyme induction, or toxic effects like DNA strand breakage. Previous work with biomarkers in cnidarians showed stimulation of lysosomal hydrolase activity in *Campanularia flexuosa* exposed to copper and the induction of heat shock proteins in the scyphozoan *Aurelia* sp. (Moore and Stebbing, 1976; Black and Bloom, 1984). In this study an adaptive response was measured: the response of catalase activity to copper-induced oxidative stress.

Catalase is a hematin-containing enzyme that catalyzes the conversion of hydrogen peroxide to water and molecular oxygen ($2H_2O_2$ ----> $2H_2O + O_2$). Catalase is mostly found in mitochondria and peroxisomes, and scavenges H_2O_2 generated during electron transport chain reactions, and fatty acid oxidation, respectively. Removal of hydrogen peroxide and peroxide side groups is catalyzed by both ascorbate peroxidase and glutathione peroxidase, and these two enzymes share the hydrogen peroxide dismutation function of catalase.

An increase in activated oxygen species in algal-host symbionts due to photosynthetically induced hyperoxic conditions has been demonstrated. Dykens and Shick (1982) showed increased partial pressures of O_2 in photosynthetically active *A. elegantissima* . Superoxide dismutase activity was positively related to chlorophyll content (an indication of algal biomass). Following this, Dykens' group discovered direct evidence of light dependent oxy-radical formation in *A. elegantissima* endosymbiotic zooxanthellae and host tissue (Dykens el at., 1992). Hydroxyl radical and superoxide anion

production occurred in aposymbiotic (algae-free) anemones as well, possibly indicating radical production from direct photoexcitations.

Catalase activity was induced by increasing photosynthetically active radiation and UV radiation exposures to intact symbioses and isolated zooxanthellae from the anthozoan *Aiptasia pallida* (Lesser and Shick, 1989a; Lesser, 1989). This was also the case in reciprocal transplant experiments between low and high light intensities, although the effect was weaker (Lesser and Shick, 1989a). Shick and Dykens (1985) found catalase activity to correspond with chlorophyll-related SOD activity in their survey of Great Barrier Reef symbiotic invertebrates. Dykens (1984) showed a direct relationship between catalase activity and chlorophyll content in *A. elegantissima*. Catalase was photoinactivated by high light intensities (400-450 $\mu E \cdot m^{-2} \cdot s^{-1}$ PAR for 4 hours) in *Aiptasia pallida*, possibly due to damage caused by hydroxyl radicals (Tapley, 1989). In view of the mechanistic toxicology of copper and the abundance of activated oxygen species in *A. elegantissima*, it is expected that the deleterious effects of oxy-radicals would be potentiated by copper and that subsequent hydrogen peroxide production may induce catalase activity.

Among copper-exposed biological systems examined, catalase may be either induced or depressed. Induction may occur in response to influxes of hydrogen peroxide, whereas depression may be caused by site-specific radical damage. Catalase inhibited damage to mammalian DNA in *in-vitro* systems containing copper and hydrogen peroxide (Aruoma et al., 1991). Catalase activity was depressed in copper-exposed common carp (*Cyprinus carpio morpha*) (Radi and Matkovics, 1988). Glutathione peroxidase activity was, however, stimulated. Copper was strongly inhibitory *in-vitro* to catalase from tissues of the fish *Sarotherodon mossambicus* (Singh and Sivalingham, 1982). *Fundulus heteroclitus* (mummichog) exposed to copper for 96 hours showed depression of catalase in both whole animal and *in vitro* exposures (Jackim, 1974).

OBJECTIVE OF STUDY

This study was designed to evaluate the utility of *A. elegantissima* as a candidate species for toxicity biomonitoring. A deficiency of field biomonitoring programs has been the under-utilization of biological effects monitoring and inattention to rocky shoreline communities. Where effects monitoring has been undertaken, only several species have been examined. Monitoring with *Mytilus edulis* has been hampered by their intermittent reproduction, the interference of the shell to pollutant uptake, and by their tendency to cease pumping when exposed to some toxicants. These problems may be overcome through the development of more suitable monitoring systems. *Anthopleura elegantissima*, and other cnidarians, may prove to be valuable organisms for environmental assessment efforts.

As a first tier evaluation, the first objective was to determine appropriate sublethal dosages through an acute toxicity rangefinding exposure. The second objective was the measurement of dose responses of a suite of measurement endpoints that take advantage of the biology of *A. elegantissima*. The last objective was an evaluation of the experimental results with respect to biomonitoring efficacy.

The results represent an initial evaluation of *A. elegantissima* as candidate species for toxicity studies. Chronic effects were measured at levels far below acutely lethal concentrations and bioaccumulation was linear with increasing dose. While effects were not in a range commensurate with ambient copper concentrations at polluted marine environments, the results do show that *A. elegantissima* is amenable to toxicity testing and further experimentation with other toxicants or measurement endpoints may show suitable sensitivity.

METHODS AND MATERIALS

Seawater, Toxicant, Water Quality Instrumentation and Glassware

Seawater used in the rangefinding and sublethal tests came from the open-circuit seawater system at Shannon Point Marine Center at, Washington. Seawater from this system was filtered through 5-micron spun fiberglass filters and trickled through activated carbon before storage in acid washed 20 liter high-density polyethylene carboys. The filtered sea water (FSW) was allowed to age for at least 2 weeks at 10°C before use to allow the oxidation of dissolved organic matter.

Cupric sulfate pentahydrate, $CuSO_4 \cdot 5H_2O$ (Sigma, lot 38F-0527), was used for both the acute exposure and the sublethal exposure experiments. A stock toxicant solution in filtered sea water was made at the beginning of the experiment and kept at 5° C. Serial dilutions were made at each test media renewal. Reported test group toxicant concentrations are calculated from the dilution series.

All glassware and materials in contact with the anemones, toxicant, or dilution water, with the exception of the pH and dissolved oxygen probes, were washed with phosphate-free laboratory soap, rinsed 5X with tap water, 2X with tap distilled water, acid washed with 10% HNO_3 solution, rinsed 7X with deionized-distilled water, and allowed to air dry. Probes (D.O. and pH) were rinsed with copious amounts of tap distilled and distilled-deionized water between sampling different treatment groups. Measurements with these probes were made in order of increasing toxicant concentration to avoid carryover of copper. During toxicant exposure periods, pH and dissolved oxygen were measured at every media renewal. Dissolved oxygen was measured with a Yellow Springs Instruments, Inc. combination D.O.-temperature-salinity meter, model 57, with a Clark electrode. The meter was calibrated before each use by the water saturated air method corrected to salinity, temperature and barometric pressure. An Orion, Inc. model 231 ionanalyzer with a Ross combination electrode was used to measure pH. A two point calibration at pH 4 and pH 7 with buffers at 10°C was conducted before each use.

Acute Toxicity Rangefinding Test

Zooxanthellate anemones used in the rangefinding acute toxicity phase of the experiment were collected from one large tidepool at Sares Head on the western shore of Whidbey Island, Island County, Washington, U.S.A (Lat. 48°26'W; Long. 122°41'N). Sares Head is a conglomerate and boulder outcrop located approximately one-half mile north of Deception Pass (Fig. 1). The tidepool is located near the mean low water level and the anemones were collected during a -1 foot tide. A total of 65 anemones were collected on August 9, 1991 using sharpened wooden medical tongue depressors to dislodge the pedal disc from the substrate. An attempt was made to collect clonemates in the sample by collecting anemones from one contiguous aggregation. The sample was transported within two hours of collection in two acid washed 5 gallon pvc buckets containing freshly collected seawater to the laboratory at Western Washington University. Upon arrival, intact anemones were placed individually in numbered, 250 ml glass beakers containing 150 ml filtered seawater.

The anemones were acclimated to laboratory conditions for 33 days at 10°C on a 12:12 photoperiod in a laboratory incubator. Light intensity at tray level under cool fluorescent lights was approximately 165 μEinsteins m^{-2} s^{-1} measured with a Licor quantum photometer model LI-185. Beakers were rotated to equalize light exposures, and incubator temperature was recorded at every culture water renewal. Anemones were fed chopped fresh *Mytilus edulis* tissue approximately 12 hours prior to media renewal; at the beginning of the light period. Wet weight of mussel tissue portions fed to each anemone was measured to the nearest 0.1 gram on a top-loading balance. Mussel tissue was offered to the anemones with forceps and direct contact of the food to the inner tentacles. Approximately 12 hours after feeding the sides of the beakers were cleaned and culture water was renewed. Debris and mucus was removed from the anemones and beakers were cleaned with an acid washed rubber policeman and FSW in a Nalgene squirt bottle. Anemones with damaged pedal disks and anemones that did not feed were culled during the acclimation period.

To explore tentacle regeneration, surgical scissors were used to clip one inner whorl tentacle from anemones anesthetized with 4.75 mg/l MgCl in FSW. Tentacles were clipped 2 days prior to beginning the exposure period. Maximum time in the anesthetic solution was two minutes, after which

anemones were placed in fresh seawater. Subsequent attempts to measure tentacle length with a hand-held scale, an ocular micrometer, photomicroscopy and a micromanipulator were hampered by anemone movements and this effort was abandoned.

The dosing period began by randomly assigning healthy anemones to treatment levels. Serial dilutions of a 7.821 mg/l $CuSO_4$ stock solution in FSW were made to the log_{10} series: 1990.74, 199.07, 19.91, and 1.99 ug/l CuSO4 (31.33 µM, 3.133 µM, 0.3133 µM, and 0.0313 µM Cu). Treatment groups consisted of 9 anemones per copper concentration and included 9 anemones in a zero dose reference group exposed to FSW only. The timetable for experimental procedures is illustrated in Figure 2.

Assay maintenance during the exposure period was the same as during the acclimation period. Dissolved oxygen and pH were measured immediately prior to media renewal. Response to gentle prodding, retraction, and food aversion were noted at media renewal. Criteria for food aversion and anemone retraction were the presence of intact mussel tissue and the covering of the polyp mouth by tentacles, respectively, at media renewal. Egested pellets were not counted as food aversion incidence. Anemones that failed to respond to gentle prodding with any retraction movement for two consecutive days were removed as putative mortalities. The test was terminated after 28 days of exposure.

I determined appropriate exposure conditions during this experiment. For example, dissolved oxygen concentrations reached very low levels if the interval between feeding and test solution renewal extended beyond 8 hours, or if chopped mussel rations were greater than approximately 0.4 grams wet weight. The lethal copper concentration and anemone maintenance procedures determined during the rangefinding test were utilized in the sublethal exposure experiment.

Figure 1. ***Anthopleura elegantissima*** **Collection Site Vicinity Map**

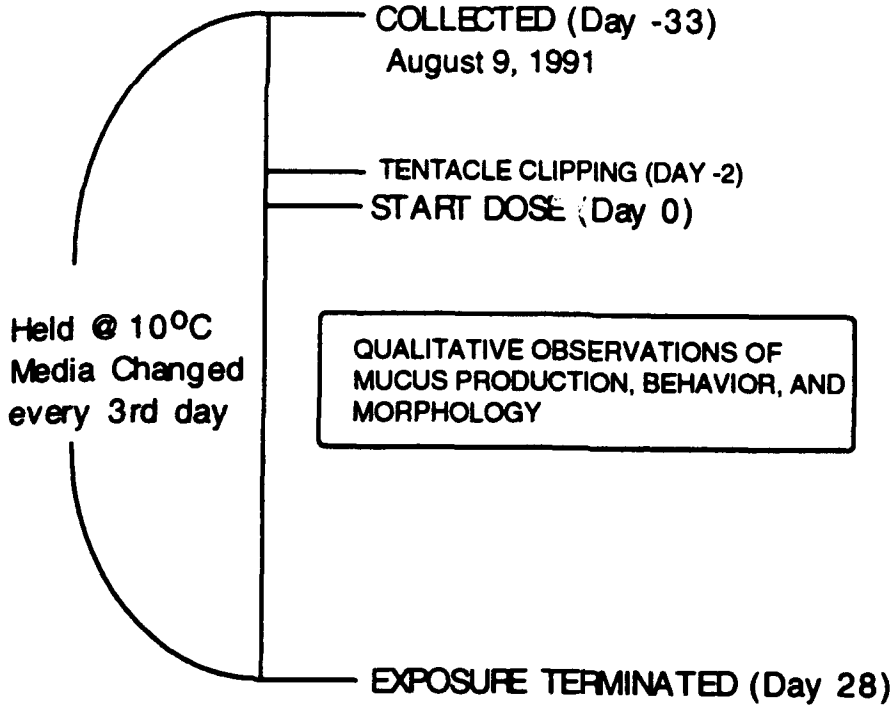


Figure 2. Timeline for Collection and Exposure: Acute Toxicity Rangefinding Test.

Sublethal Test

Seventy-five zooxanthellate anemones were collected on December 3, 1991 during a zero-foot tide from an exposed sandstone outcrop at the western shore of Samish Bay, Skagit County, Washington, U.S.A. (Lat. 48°38'W; Long. 122°29'N). Samish Bay is a large, relatively unprotected, sand and boulder substrate bay (Fig. 1). The collection area was a boulder tidepool complex approximately +1 foot above mean low water level. Anemones were transported in two 5 gallon pvc buckets within one hour of collection to the laboratory at Western Washington University where they were immediately placed individually in 250 ml glass beakers containing 150 ml FSW in a 10°C. incubator.

The day after collection, the anemones were transported in two 5 gallon pvc buckets on ice containing FSW to a running seawater table at Shannon Point Marine Center where they were kept, unfed, under natural light to remove sand and shells adhering to the anemones. They were returned to Western Washington University 9 days after collection and placed individually in numbered 250 ml glass beakers in a 10°C. incubator with a 12:12 photoperiod under Sylvania Gro-Lux fluorescent lights. Light intensity at tray level averaged 140 μEinsteins m^{-2} s^{-1}. Specimen handling procedures were identical to the acute test. The timetable for sample handling and the exposure period is illustrated in Figure 3.

Baseline reduced weight measurements were taken 2 days prior to commencing the exposure period. Reduced weight was measured by weighing below the pan of a Mettler AE163 analytical balance (Muscatine 1961). Dislodged anemones were suspended in FSW by a hooked 8cmX0.25mm constantan wire inserted into the actinopharynx and displacement mass was determined. The balance was internally calibrated, and FSW temperatures were recorded after every 7 anemones were weighed.

The sublethal exposure period commenced 22 days after sample collection. In addition to a zero dose reference group, a toxicant stock solution was serially diluted to prepare nominal 250, 175, 100, and 25 ug/l Cu treatment concentrations (3.934 μM, 2.754 μM, 1.574 μM, and 0.3934 μM Cu). Sixty-five healthy anemones were randomly assigned to 5 treatment levels, each with 13 replicates individually exposed to 150 ml treatment media in 250 ml glass beakers. The media renewal, feeding, behavior data collection, beaker rotation and water quality measurement regimen was performed as in the acute test

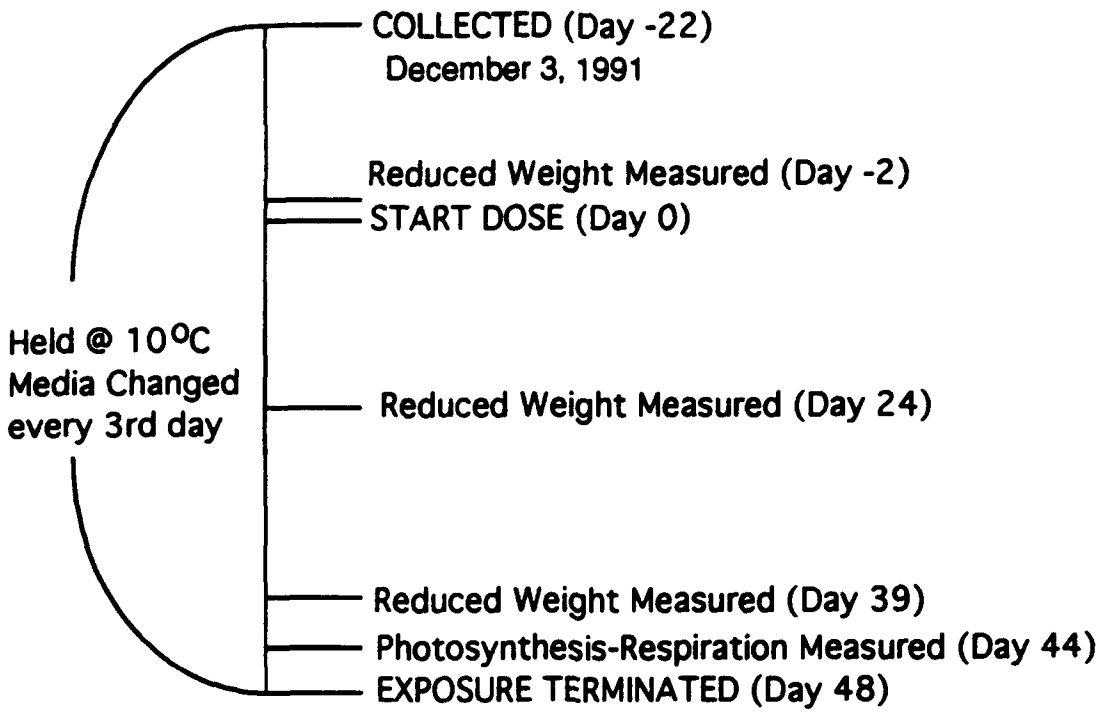


Figure 3. Timeline for Collection and Exposure: Sublethal Test.

described above, except a maximum of 8 hours was allowed between feeding and media renewal. Midway through the exposure period, on day 24, and, again near the end of the exposure period on day 39, the anemones were dislodged from their beakers with sharpened tongue depressors and reduced weight was measured in FSW. Anemones were not fed for 3 days prior to reduced weight measurement. Tentacle retraction and feeding aversion were determined as in the acute test. Behavior data from 11 media renewals were tallied to give summations of aversion and retraction behavior occurrences and table values indicated are the group means.

The toxicant exposure period was terminated the morning of day 48. Anemones were dislodged from their beakers, rinsed 2X with FSW, blotted dry, placed individually in whirlpack bags and transported on ice for storage at -70°C.

Photosynthesis and Respiration

On days 41, 42 & 44 photosynthesis and respiration rates of anemones were measured by measuring dissolved oxygen flux in the light and in the dark in sealed flasks containing exposed and control group anemones. Four randomly selected anemones from the control, 100, and 250 µg/l treatment groups were placed in 125 ml Erlenmeyer flasks containing filtered, undosed, seawater and a micro-stir bar. Photosynthesis and respiration of each anemone was measured singly. Clark type dissolved oxygen electrodes were inserted into each flask. Care was taken to remove all visible bubbles of air from each flask and the probes were sealed in place with Büchner funnel gaskets and parafilm. The probes were calibrated by the water saturated air method. Probes were connected to an Endeco type 1125 pulsed D.O. sensor controller, and temperature and oxygen data were recorded every 2 minutes with a computerized data collection system. A circulating water bath was used to maintain constant (10°±4 C) temperature.

Each flask was placed equidistant from fluorescent light banks and directly above a magnetic stir bar motor to ensure stirred conditions within the flasks. Light intensity averaged 230 µEinsteins m^{-2} s^{-1} under water at a distance of ~5 cm from the fluorescent bulb to the flask wall. The system was operated for 30 minutes prior to placement of the anemones to allow the probes to equilibrate. Anemones were allowed to acclimate for 30 minutes under light, then oxygen concentration was monitored for 30 minutes under light, 30

minutes in darkness, and then for 30 minutes under light again. At the conclusion of the measurements, the anemones were returned to their beakers and exposure regimen. Photosythesis and respiration rates were calculated from the slope of the D.O. flux, and the volume of the flask containing the anemone and probe. Final data are expressed as $\mu g\ O_2\ g^{-1}$ anemone dry weight h^{-1}. Poor temperature control, and air bubbles rendered data from days 41 and 42 of poor quality and only data from day 44 are included herein.

Sample Preparation and Tissue Copper Analysis

Frozen anemones were individually weighed to the nearest 0.1 mg then homogenized in a total volume of anemone plus clean seawater of 20 ml for 45 seconds at mid-speed in a Virtis model 20 electric tissue homogenizer. Four 1 ml aliquots of homogenate were placed in new plastic 1.5 ml Eppendorf tubes and were frozen at -20°C. The remaining anemone homogenates were placed individually in pre-muffled, pre-weighed ceramic crucibles and were dried at 105°C until a constant dry weight was obtained. .

Sample preparation for atomic absorption spectrophotometric analysis of copper in anemone tissues followed the dry ashing procedure (AOAC, 1980). Crucibles containing dried homogenate were muffled at 450°C for 12 hours. Four ml of reagent grade concentrated nitric acid was added to each cooled crucible and then slowly evaporated off over a laboratory hot plate. Crucibles were subsequently muffled at 450°C for one hour and cooled. A 1N solution of reagent grade HCl in distilled, deionized water was used to dissolve the ash. The crucibles were rinsed with 3 aliquots of the acid solution and the dissolved sample was brought to 25 ml in a class A volumetric flask. The acidified samples were stored at 4°C in 60 ml Nalgene bottles until analysis.

A Perkin Elmer atomic absorption spectrophotometer model 560 with oxidizing air-acetylene flame was used to measure sample absorption by direct aspiration at 324.8 nm wavelength, and 0.7 nm slit width (0.077 mg/l sensitivity, linear range up to 5 mg/l Cu) (Perkin Elmer, 1982). Sample absorption was measured relative to a serial dilution of a dry ashed copper standard in FSW, and a dry ashed FSW blank prepared with class A glassware. Samples and standards at room temperature were measured in triplicates of 5 second processor derived averages. The spectrophotometer was recalibrated every 10 samples. Salt buildup required periodic cleaning of the burner head and

recalibration. Tissue copper levels were calculated using the standard curve and are expressed as μg Cu/g anemone dry weight.

Zooxanthellae Density and Cell Division

One ml aliquots of anemone homogenate in Eppendorf tubes were thawed on ice, and vortexed 3X for 20 seconds. Remaining tissue clumps were homogenized by adding 0.2 ml of homogenate to approximately 1 ml of filtered seawater in a 2 ml Wheaton glass hand-held tissue grinder, and grinding for 25 strokes. The grinder contents were quantitatively diluted with filtered seawater, vortexed, and cells observed at 400X on an Improved Neubauer hemacytometer grid. Homogenate dilutions were made to yield approximately $1X10^6$ cells/ml. A total of 6 subsample counts were made for each homogenate sample. Cells undergoing cytokinesis were noted if a division furrow was observed. Data are expressed as cells per μg animal supernatant protein.

Supernatant Protein Concentration

To provide a more accurate expression of per unit animal zooxanthellae density and catalase activity, the protein concentration of the homogenate supernatant was determined. Protein was quantified relative to a bovine serum albumin standard by the protein-dye binding method of Bradford (1976). The calibration series was made by serial dilution of a 1.0 mg/ml distilled deionized water-bovine serum albumin stock solution (BSA Sigma, lot 88FO790). The Bradford reagent was made by dissolving 143.0 mg of 70% Coomassie Blue G dye in 100 ml of absolute ethanol, to which 100 ml of 85% phosphoric acid was then added. This solution was brought to 1 liter with distilled deionized water and then filtered through a Whatman #1 filter. The Bradford reagent was stored at 4°C and brought to room temperature before assay.

Thawed 1.0 ml aliquots of homogenate in Eppendorf tubes were vortexed 3X for 20 seconds and centrifuged at 11,000 rpm in a Sorvall centrifuge model 24S for 5 minutes at room temperature. Three 0.1 ml subsamples of the supernatant were individually diluted with distilled-deionized water, vortexed, and the Bradford reagent was added. The final mixture was briefly vortexed and the absorbence at 595 nm was recorded exactly [illegible] minutes after the addition of the dye reagent with a Perkin Elmer Lambda 3B uv/visible spectrophotometer. Sample protein concentrations were calculated from regressions of BSA standard dilution absorbence.

Catalase Activity

Depletion of hydrogen peroxide by the first order reaction $2H_2O_2 \dashrightarrow 2H_2O+O_2$ (literature rate constant $k=10^7$ l X mol^{-1} X s^{-1}) was monitored to quantify catalase activity in supernatants of anemone homogenates by the method of Beers and Sizer (1952), with modifications contributed by David Tapley (pers. communication). This method has the advantage of determining the concentrations of enzyme and substrate, which is a necessity due to the first order nature of the reaction.

Catalase activity was measured by continuously monitoring the decrease in the absorbance of H_2O_2 at 240 nm with an IBM model 9420 UV-Visible spectrophotometer. Eppendorf tubes containing 1.0 ml aliquots of frozen anemone homogenate were thawed at 5°C, vortexed 3X for 20s, and centrifuged at 11,000 rpm for 5 minutes in a Sorvall 24s centrifuge at 5°C. The substrate was a freshly made 4.4%v/v solution of ~30% reagent grade H_2O_2 dissolved in 50mM phosphate buffer pH7. Strength of the substrate solution was checked by adding 100 µl to 3 ml buffer and measuring the absorbance at 240 nm. Substrate solution strength was adjusted to between 0.500 and 0.550 A_{240} prior to beginning the assay.

Between 40 and 120 µl of supernatant was added to 3 ml of 50mM pH 7 phosphate buffer at room temperature (~22° C) in a quartz cuvette. Background absorbance of the supernatant-buffer solution was measured to ensure turbidity remained below 1 absorbance unit. The reaction was initiated by adding 100 µl substrate solution. The decline in absorbance of the sample relative to a quartz reference cuvette containing buffer only was continuously monitored on a chart recorder until the absorbance decreased by 0.05 units. Catalase measurements were made in triplicates from each supernatant sample. Catalase activity was calculated from the change in absorbance and is expressed as µMole H_2O_2 s^{-1} mg^{-1} protein.

STATISTICAL ANALYSIS

Data were tested for homoscedasticity by Shapiro-Wilk's test, and normal distribution by Bartlett's test using the PC-based statistical package Toxstat release 3.2 (Toxstat, 1990). Data that could not be transformed to meet the above assumptions of analysis of variance were further analyzed for significance with the Toxstat hosted Kruskal-Wallis nonparametric test. Descriptive statistics, correlation, regression, and one way analysis of variance significance testing were calculated using the SPSS Release 4.0 statistical program on the Western Washington University VMS computer (SPSS, Inc., 1992). Non-metric clustering and association analyses were calculated with the RIFFLE program (Matthews and Heame, 1991). All significance levels were set *a-priori* at 5%. The rangefinding test EC_{50} values are graphical estimates as lack of partial mortality in more than one treatment group did not allow calculation of confidence limits (Webber, 1991).

RESULTS

Anthopleura elegantissima exhibited sublethal responses well below acutely toxic doses of copper. The lowest observed statistically significant concentration for a biological effect was 175 µg/L for percent weight gain in the second half of the exposure period. Behavioral endpoints were inhibited and zooxanthellae reproduction was stimulated in the 250 µg/L treatment when compared to the control group. Copper was bioaccumulated without apparent regulation of uptake. Each of the endpoint responses are addressed individually below.

RANGEFINDING TEST

After 24 days exposure to copper sulfate, total cumulative mortality was 78% in the highest concentration treatment group tested (2000 µg/L copper). None of the other treatment groups sustained mortality. The graphical EC_{50} estimate was 1350 µg/l. Mucus production, tentacle retraction, and food aversion were more pronounced with increasing dose. Mucus production was extreme in the 2000 µg/L group and mucus was removed from the anemones at each media renewal. Mid and high dose (200 and 2000 µg/L Cu) anemones showed stress at media changes by increased swelling and extension of the column. Squash mounts of clipped tentacles viewed under magnification indicated the presence of unidentified green-brown granules in the mid and high dose groups.

SUBLETHAL EXPOSURE

Anemone Growth

Initial reduced weight was not significantly different between the groups (initial mean reduced weight=0.1163 g; sd=0.0486). Percentage difference was calculated from reduced weight data to yield calculated percent gain data for the first half of the exposure period, the second half, and percent gain over the entire exposure period. Group averages, standard deviations and ANOVA significance are shown in Table 1. The percent weight gain dose response for the first and second intervals of the exposure period is shown in Figure 4. Weight gain was not significantly different in any of the groups at the mid point of the exposure period. Weight gain was affected by treatment concentration in

the second half of the exposure period, with severe inhibition in the 175 and 250 µg/L groups. The concentration response for overall percent weight gain showed statistically significant inhibition of weight gain in the 175 and 250 µg/L groups relative to the control group (Fig. 5).

Behavior

Statistically significant behavioral effects were found with increasing dose. Kruskal-Wallis nonparametric ANOVA showed a significant increase in tentacle retraction behavior in the highest dose tested (250 µg/L) compared to the 25 µg/L and control groups (Table 2). Likewise, feeding aversion was significantly greater in the 250 µg/L group compared to the 25 µg/L and control groups. Feeding aversion responded to treatment concentration in a biphasic manner with weak (statistically non-significant) stimulation of feeding at the 25 and 100 µg/L levels. Feeding aversion in the 175 µg/L group was approximately equal to the control group response, while aversion was pronounced in the 250 µg/L group (Fig. 6a). There was no bimodality in the tentacle retraction behavioral response; increasing test concentration yielded more frequent tentacle retraction (Fig. 6b). Tentacle retraction and feeding aversion were greatest at the middle of the exposure period.

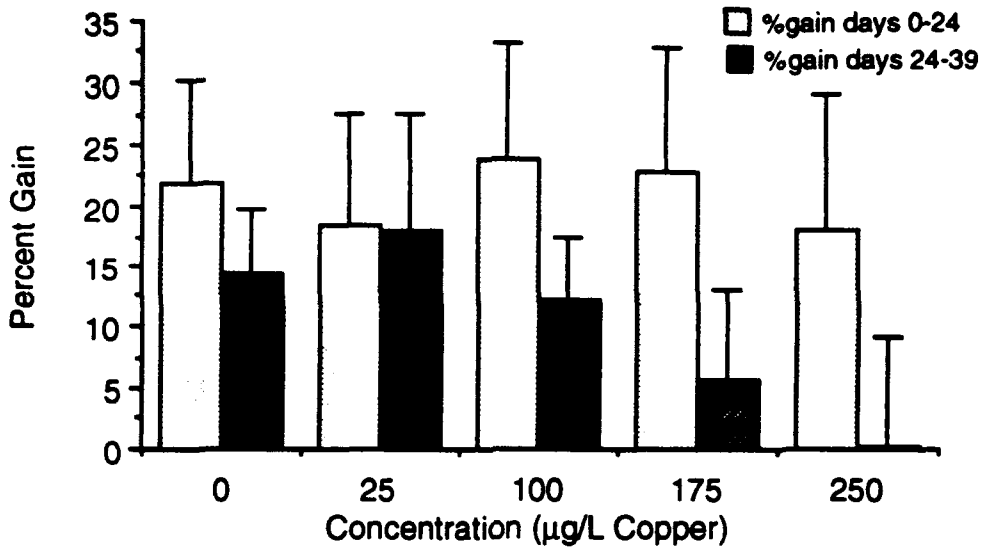


Figure 4. Percent Gain in Reduced Weight During the First 24 days and During the Last 15 Days of Exposure to Copper. Error bars represent standard deviation (n=13).

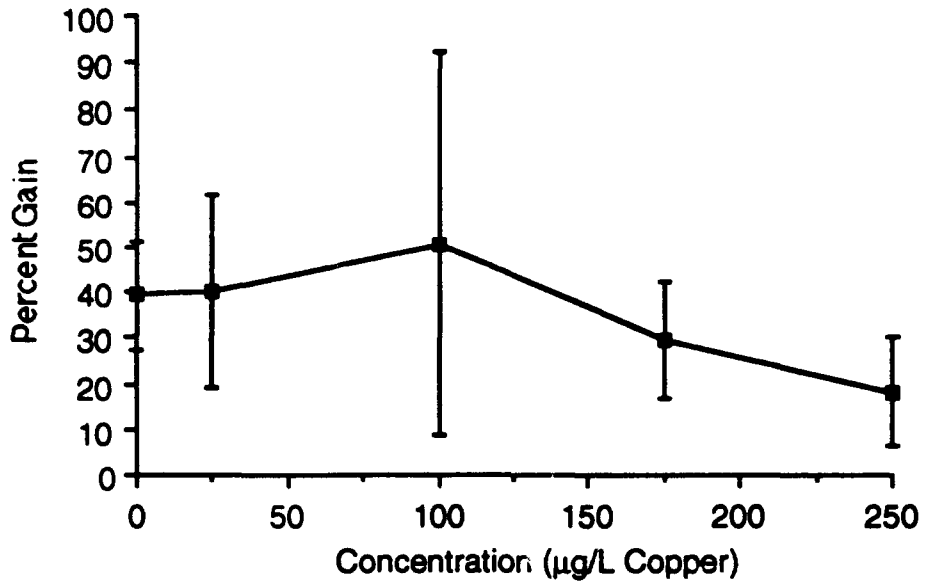


Figure 5. Percent Gain in Reduced Weight of Anemones with Dose of Copper Over the Entire Exposure Period. Error bars represent standard deviation (n=13).

Concentration µg/L Copper	**After 24 Days** percent weight gain (SD)	ANOVA SIGNIFICANCE *Fprob.=0.421:Fcalc=0.989*
0	21.9 (8.3)	*no sig. diff.*
25	18.4 (9.3)	
100	23.9 (9.4)	
175	22.9 (9.9)	
250	18.1 (10.9)	

Concentration µg/L Copper	**Between Days 24-39** percent weight gain (SD)	ANOVA SIGNIFICANCE Fprob.=0.000:Fcalc=11.318
		250 175 100 0 25
0	14.3 (5.5)	250
25	18.1 (9.5)	175
100	12.2 (7.5)	100 * *
175	5.6 (7.5)	0 * *
250	0.2 (9.0)	25 * *

Concentration µg/L Copper	**Between Days 0-39** percent weight gain (SD)	ANOVA SIGNIFICANCE Fprob.=0.0013:Fcalc=5.110
		250 175 0 25 100
0	39.4 (12.0)	250
25	40.4 (21.1)	175
100	50.7 (41.9)	0 *
175	29.7 (12.9)	25 *
250	18.2 (12.0)	100 *

Table 1. Percent Gain in Reduced Weight Results. The tables indicate treatment group means, standard deviations and Duncan's multiple range test results for the first half (days 0-24), second half (days 24-39), and from the start to the finish of the exposure period (days 0-39), respectively (n=13).

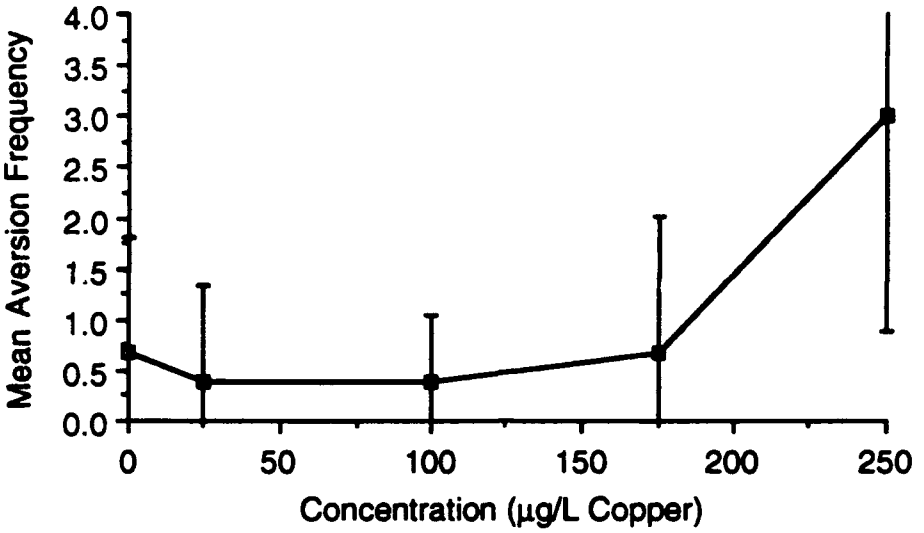


Figure 6a. Feeding Aversion Frequency Dose Response. Error bars represent standard deviation (n=13).

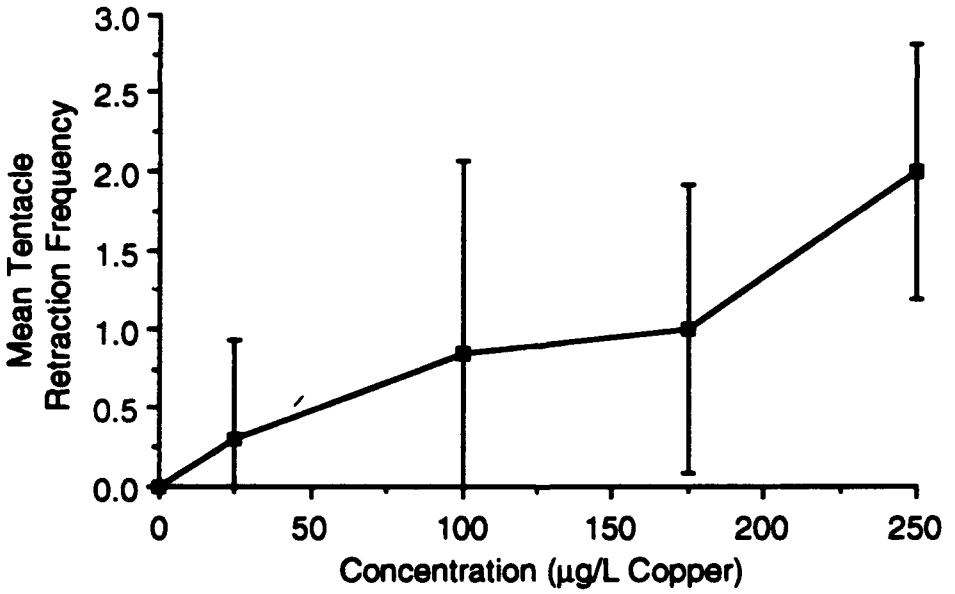


Figure 6b. Tentacle Retraction Frequency Dose Response. Error bars represent standard deviation (n=13).

Concentration µg/L Copper	**FEEDING AVERSION** mean after 11 media changes	ANOVA SIGNIFICANCE KRUSKAL-WALLIS				
			0	25	100	175	250
0	0.69 (1.11)	0					
25	0.39 (0.96)	25					
100	0.39 (0.65)	100					
175	0.69 (1.32)	175					
250	3.00 (2.12)	250	*	*			

Concentration µg/L Copper	**TENTACLE RETRACTION** mean after 11 media changes	ANOVA SIGNIFICANCE KRUSKAL-WALLIS				
			0	25	100	175	250
0	0.00 (0.00)	0					
25	0.31 (0.63)	25					
100	0.85 (1.21)	100					
175	1.00 (0.91)	175					
250	2.00 (0.82)	250	*	*			

Table 2. Anemone Behavior Frequency. Feeding aversion and tentacle retraction frequency treatment means, standard deviations, and Kruskal Wallis non-parametric ANOVA significance (n=13).

Photosynthesis and Respiration

No statistically significant gross specific photosynthesis or specific respiration rate, or P_{gross}/R ratio concentration responses were observed between the 0, 100, and 250 µg/L groups after 44 days exposure to copper. Group averages and standard deviations are listed in Table 3. Anemones from the 25 and 175 µg/L groups were not tested.

Copper Bioaccumulation

Copper levels in whole anemone homogenate increased linearly with concentration, with no apparent regulation of copper uptake at low dosage (Tissue Burden =7.22(SE=1.36)+0.216(SE=0.009)XDose) (R^2=0.893; Fcalc=526.1; Fprob=0.000) (Fig. 7). Analysis of variance showed each group mean tissue copper level to be significantly different from each other except for the 25 µg/L and control groups (Table 4).

Zooxanthellae Density and Cell Division

Zooxanthellae density was essentially constant among the treatment groups, with no statistically significant differences detected by analysis of variance. Both the number of zooxanthellae cells at the cytokinesis stage (mitotic) and percent mitotic index was significantly elevated in the 250 µg/L group relative to the control, 25, and 100 µg/L groups (Table 5). The mitotic index concentration response is shown in Figure 8.

Catalase Activity

Analysis of variance did not reveal any significant catalase activity differences between any of the dose groups. Group means and standard deviations are in Table 6.

Conc. µg/L Copper	PHOTOSYNTHESIS (GROSS) µg O2/g dry wt./hour	RESPIRATION µg O2/g dry wt./hour	P:R	n=
0	189.4 (84.4)	252.6 (100.8)	0.74 (0.81)	4
25	-	-	-	
100	157.0 (35.0)	184.4 (19.3)	0.85 (0.13)	3
175	-	-	-	
250	121.9 (45.6)	169.7 (58.7)	0.71 (0.12)	4

Table 3. Photosynthesis, Respiration and P_{gross}:R Ratio Results. Table shows treatment means and standard deviations at day 44 of the exposure period. No statistically significant differences were detected.

Concentration µg/L Copper	**COPPER** µg Cu/g dry wt.	ANOVA SIGNIFICANCE Fprob=0.00:Fcalc=132.7
		0 25 100 175 250
0	8.1 (4.4)	0
25	12.8 (4.8)	25
100	26.1 (6.8)	100 * *
175	47.0 (9.1)	175 * * *
250	61.0 (8.8)	250 * * * *

Table 4. Copper Bioaccumulation in Anemone Tissue After 48 Days of Exposure to Copper. Table indicates treatment means, standard deviations and Duncan's multiple range test results (n=13).

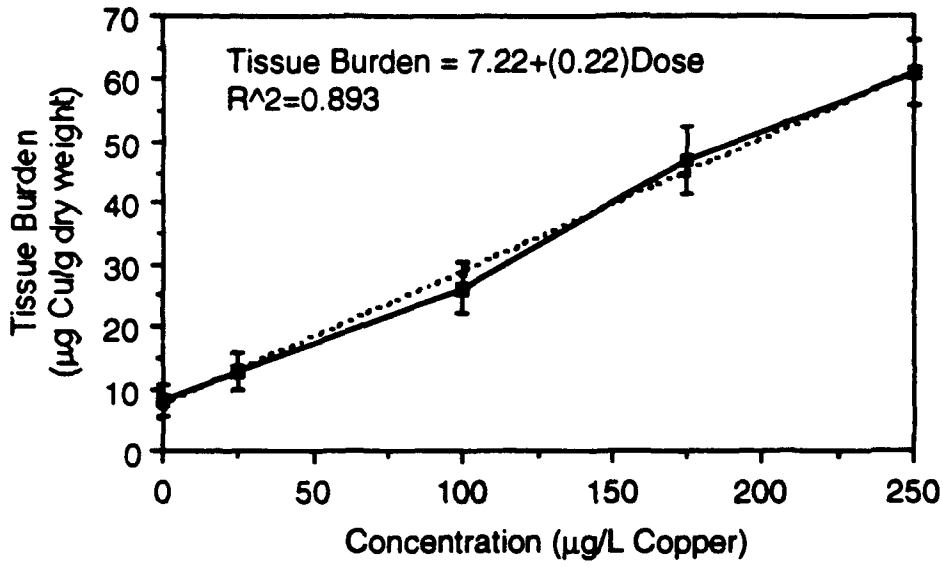


Figure 7. Tissue Copper Uptake After 48 Days of Exposure to Copper. Error bars represent standard deviation. Least squares regression indicated by hatched line (n=13).

Concentration µg/L Copper	**ZOOXANTHELLAE** cells/ µg protein (SD)	ANOVA SIGNIFICANCE *Fprob.=0.640:Fcalc=0.634*
0	$6.895X10^4$ ($1.946X10^4$)	*no sig. diff.*
25	$6.799X10^4$ ($1.478X10^4$)	
100	$6.779X10^4$ ($1.380X10^4$)	
175	$6.180X10^4$ ($1.405X10^4$)	
250	$7.096X10^4$ ($1.475X10^4$)	

Concentration µg/L Copper	**DIVIDING ZOOX.** cells/ µg protein (SD)	ANOVA SIGNIFICANCE Fprob.=0.044:Fcalc=2.615
		100 25 0 175 250
0	1003 (663)	100
25	935 (517)	25
100	914 (336)	0
175	1088 (527)	175
250	1472 (442)	250 * * *
Concentration µg/L Copper	**MITOTIC INDEX** percent (SD)	ANOVA SIGNIFICANCE Fprob.=0.0496:Fcalc=2.531
		100 25 0 175 250
0	1.5 (0.7)	100
25	1.4 (0.8)	25
100	1.4 (0.5)	0
175	1.7 (0.8)	175
250	2.1 (0.6)	250 * * *

Table 5. Zooxantheliae Density and Cell Division Results. Table indicates treatment group means, standard deviations and Duncan's multiple range test results (n=13).

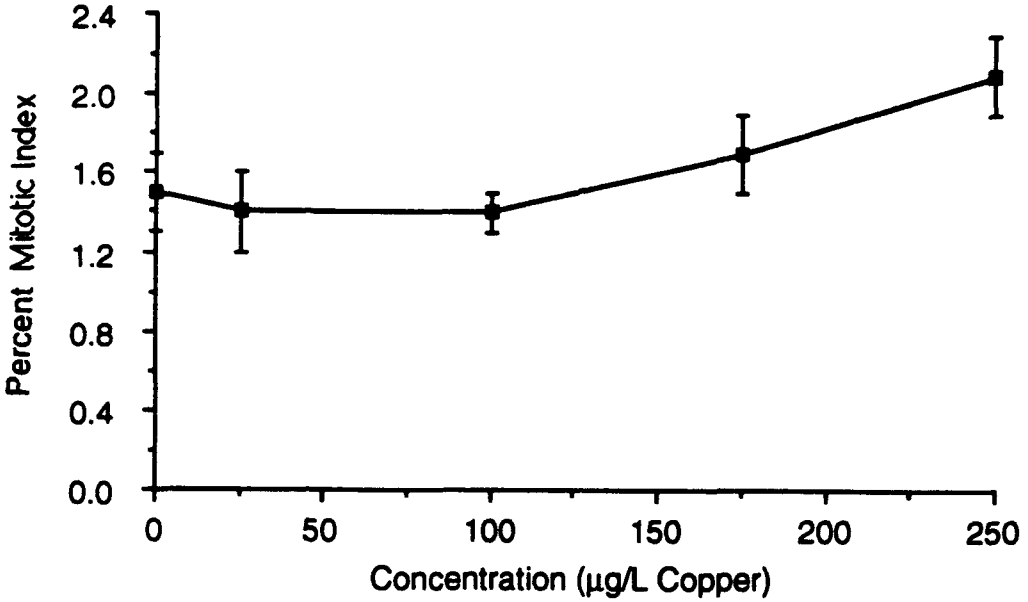


Figure 8. Mitotic Index of Zooxanthellae in *A. elegantissima* After 48 Days Exposure to Copper. Error bars represent standard deviation (n=13).

Concentration	**CATALASE**	ANOVA SIGNIFICANCE
µg/L Copper	µMoles H2O2/min./mg prot.	*Fprob.=0.111*
0	18.15 (7.7)	*Fcalc=1.969*
25	23.57 (6.58)	
100	24.49 (7.48)	*no sig. diff.*
175	22.02 (4.98)	
250	18.25 (10.27)	

Table 6. Catalase Activity in the Anemone *A. elegantissima* (animal fraction). Mean catalase activity after 48 days of exposure to copper. Standard deviations are in parentheses (n=13).

Multivariate Evaluations

Multivariate evaluations included MANOVA, and non-metric clustering. Multivariate analysis of variance indicated a significant difference between the treatment groups (Pillais F=1.659; F_{prob}<0.001). This supports, based on the whole data set (% weight gain days 0-24; % weight gain days 24-39; mean tentacle retraction frequency; mean feeding aversion frequency; tissue copper; zooxanthellae density; zooxanthellae division; gross photosynthesis; respiration; and catalase activity), the statistically significant dose responses observed in univariate ANOVA but without accrual of type 2 error.

Nonmetric clustering of the best four biological effects variables, selected by the RIFFLE algorithithm to give the largest proportional reduction in clustering error, partitioned the data set into 3 different groups that corresponded to control, medium and high dosages (Association Analysis Chi Square significance = 0.999). Similar results were obtained when the cluster was based on the best 7 variables. Variables determined to be the best clustering determinants correspond to those variables with statistically significant dose responses as determined by multiple range tests. Variables determined to be most important attributes to determining the clusters are listed in Table 7.

VARIABLE	QUALITY 4 attributes	QUALITY 7 attributes
feeding aversion	0.64	0.61
tentacle retraction	0.60	0.39
mitotic index	0.57	0.46
weight gain 2nd half	0.39	0.41
weight gain 1st half		0.52
zooxanthellae density		0.35
catalase activity		0.29
AVERAGE QUALITY:	0.55	0.43

Table 7. Non Metric Clustering Results. Variables determined to be the most significant attributes influencing clustering quality at a clustering number of 3.

DISCUSSION

The relatively high copper EC_{50} (28 days) value (1350 µg/L) determined in the rangefinding test implies that mortality would not likely be a useful endpoint in biomonitoring programs with *A. elegantissima*. Mortality, in general, has little utility in *in-situ* biomonitoring as sensitivities are often too low and response speed is usually extended compared to other sublethal responses (Van der Schalie, 1986). Tolerance of high pollutant concentrations may, however, allow sublethal effects biomonitoring to be effective in highly polluted areas.

Acutely lethal concentrations of copper were higher for *Anthopleura elegantissima* than for other cnidarians. The greater biomass (and surface to volume ratio) of anemones may reduce site-of-action toxicant concentrations as compared to the smaller hydroid species tested. With the hydroid *Eirene viridula*, Karbe (1972) noted tissue destruction within a few hours after exposure to 3000 µg/L copper and lethality within 2 days at 500 µg/L copper. The *A. elegantissima* EC_{50} estimate for copper is also much higher than median lethality results reported for *Neanthes arenaceodentata* (LC_{50}=250 µg/L; 28 days), and *Crassostrea gigas* (LC_{50}=560 µg/L; 96 hour) (Reish et al., 1976; Okazaki, 1976). Scott and Major (1972) reported threshold concentrations at 100-200 µg/L for lethality in *Mytilus edulis* exposed to copper.

Qualitative results from the rangefinding test showed increased tentacle retraction, feeding aversion, copious mucus production, and the presence of unidentified granules observable in tentacle squash mounts. *Mytilus edulis* exposed to 40 µg/L copper also exhibited granules in digestive gland cells; these were identified as lysosomal accumulations of lipofuscin granules (Viarengo et al., 1990). Many xenobiotics are toxic through free radical mechanisms, and the biochemical measurement of lipid peroxidation byproducts may be a useful biomonitoring endpoint. Similar mucus production and tentacle retraction behavior was observed by Harland and Nganro (1990) in exposures of *Anemonia viridis* to 200 µg/L copper. Mucus production may be a detoxification modality as metals and organic toxicants may be rendered less toxic through binding by exuded mucus.

The sublethal test showed growth rate inhibition as the most sensitive endpoint (Table 1). Other studies have shown cnidarian growth rate to be sensitive to copper (Karbe, 1972; Stebbing, 1976; 1979; Moore and Stebbing,

1976). Stebbing's (1979) determination of growth rate inhibition at 10 µg/L copper with *Campanularia flexuosa* is much less than the lowest observed effect concentration (LOEC) of 175 µg/L copper observed in the second half of the exposure period. The LOEC is the lowest test concentration yielding a statistically significant dose response and is therefore sensitive to experimental design. Overall weight gain was inhibited by 64% in the 250 µg/L group. In light of the drastic increase in feeding aversion and tentacle retraction, the observed growth inhibition is most likely related to nutritional deficits. Increased tentacle retraction reduces the quantity of light reaching the zooxanthellae and may lessen the zooxanthellae nutritional contribution to the anemone. The rate of decline in growth rate would increase as stored nutrients are depleted. Growth inhibition may also be attributed to increased energy allocated to mucus production. Growth rate inhibition would likely be more profound in the field as the test anemones were fed fresh mussel tissue by hand in this experiment.

Growth was weakly stimulated relative to the control group in the 100 µg/L treatment, although the response was not statistically significant. This trend is in accordance with the role of copper as a micronutrient, or possibly with a general stimulatory response to low levels of toxicants. Stebbing (1976) found transitory stimulation in *C. flexuosa* growth rates with low level exposure to cupric chloride. He posited that stimulation may be a natural homeostatic response to stress. Any stimulation due to improved nutrition would depend on a less than optimal background concentration of copper in the diluent.

As mentioned above, hydroids typically degenerate during starvation or other stressful periods. This is reversed when conditions improve. This plasticity of size is also a feature of anthozoans and may be a useful measurement in biomonitoring programs. Measurements of growth by the reduced weight method are amenable to both field and laboratory experiments.

Anthopleura elegantissima exhibited behavioral responses to copper at 250 µg/L (Table 2). Behavioral responses are, in many cases, among the most sensitive endpoints, possibly because the receptor organs are usually highly exposed to dissolved toxicants (Døving, 1991). In cnidarians, the nervous system is proximate to toxicants dissolved in the external milieu and in the coelenteron. The toxic effects of copper may be due to receptor degeneration or to nerve signal transmission interferences. Increased tentacle retraction frequency could be damaging to *A. elegantissima* through decreased prey capture efficiency, and decreased absorption of sunlight causing reduced

photosynthetic rates and a resultant decrease in translocation of nutrients from the zooxanthellae. Feeding aversion, likewise, will reduce nutrient uptake and scope for growth.

While behavioral responses have rarely been utilized in field assessments of ambient toxicity, they have been used successfully in laboratory settings (Van der Schalie, 1986). Copper exposures in the range of 20.8-25.6 µg/L caused termination of pumping in *Mytilus edulis* (Redpath and Davenport, 1988). Although significant aversion and tentacle retraction responses were not as sensitive (Table 2), *A. elegantissima* has available a more varied behavioral repertoire than *Mytilus*, and uptake of toxicants is not constrained by shell closure.

From the results of the bioaccumulation analysis (Figure 7), *A. elegantissima* does not regulate or significantly detoxify tissue copper. This lack of regulation is attractive from the standpoint of biomonitoring because it allows a clear correlation of ambient concentrations to tissue burdens. Furthermore, tissue copper was correlated with sublethal effects in this experiment. The results of the copper bioaccumulation part of the experiment are in accordance with the experiment described by Harland and Nganro (1990). While they exposed *Anemonia viridis* for only 5 days, their results indicated a mean whole anemone tissue copper burden of approximately 48 µg Cu $gram^{-1}$ dry weight in anemones exposed to 200 µg/L copper in seawater. Their results differed from mine in that they observed regulation of copper uptake at low dosage (50 µg/L) and increased zooxanthellae expulsion, whereas zooxanthellae densities were nearly constant at the end of this study.

After 3 days exposure to 40 µg/L copper, the mussel *Mytilus galloprovincialis* tissue copper concentrations were 300% of the control (Viarengo et al., 1981). After 48 days exposure, *A. elegantissima* in the 100 µg/L dose group accumulated 322% of the control group mean tissue burden. Uptake of copper by *Anemonia viridis* was approximately 1200% of the no dose control group tissue burden after 5 days exposure to 200 µg/L copper (Harland and Nganro, 1990). Thus, copper uptake rates may be higher in *M. galloprovincialis*, and *A. viridis* although the exposure period differences may be too great to facilitate the comparison. Prolonged exposure would allow time for induction of detoxification mechanisms such as metal binding proteins, and zooxanthellae expulsion. The lower accumulation reported in this experiment may represent steady-state detoxification rate influences. The two studies listed

above, represent short-term copper uptake. In light of elevated zooxanthellae division rates observed (Figure 8), the detoxification role of zooxanthellae expulsion hypothesized by Harland and Nganro may apply here as well. If zooxanthellae expulsion and growth rates both increased in response to dose, detoxification may be carried out as zooxanthellae densities remain constant across all dose groups.

Catalase activity did not show any statistically significant dose response, although there was a pattern of stimulation of catalase activity at low dosage, and weak inhibition of activity at 250 µg/L. This is in accordance with an adaptive response endpoint. Response would be diminished as the enzyme itself is inactivated by the toxicant. As oxidative stress in Cnidarians has been largely associated with high light intensities, irradiance levels in this experiment may have been inadequate to produce activated oxygen species. Light intensities in the field would be much greater. Tentacle retraction frequency increased with dose, and the resultant shading of the zooxanthellae may lessen photosynthesis rates and production of activated oxygen species. Catalase is robust to biochemical purification and analysis, making it a good biomarker from the standpoint of analytical practicality. In light of the lack of significant response in this experiment, it may not be a useful biomarker for copper intoxication.

Copper had little effect on zooxanthellae density (Table 5). Interestingly, zooxanthellae reproduction, as measured by the density of cells in cytokinesis, increased in the 250 µg/L group, relative to the control group. The observed increase in dividing zooxanthellae may be due to an indirect effect of copper on the anemone. A loss of control over the zooxanthellae by the anemone may be caused by copper intoxication. As densities were fairly constant, mitotic index dose response curves were very similar to mitotic density curves (Figure 8). Control mitotic index (M.I.) was approximately half (1.55%) of the M.I. measured in freshly collected *A. elegantissima* (2.88%) (Wilkerson et al., 1983). This may indicate that there was a decline in M.I. over time due to laboratory conditions. Stimulation of growth rates in planktonic dinoflagellates exposed to copper has been reported (Brand et al., 1986).

Attenuation of toxicity may be achieved by algae through the external release of copper complexing compounds (McKnight and Morel, 1979). Complexation afforded by both the host and the zooxanthellae may further reduce concentrations of toxic copper species. Zooxanthellae expulsion is a

common response in symbiotic cnidarians exposed to stress. Increases in zooxanthellae reproduction may offset increases in expulsion rates to allow zooxanthellae density to remain constant. An alternative view might be that there was no change in expulsion rates with dose, and the observed mitotic densities were cells "frozen" in cytokinesis. The stimulation of division may have been a transient effect occurring at the beginning of the exposure period when tissue copper levels were lower.

For a biomonitoring program, one of the endpoint criteria listed in the introduction was that the response variable should have a significant biological meaning, preferably an adverse effect. Increase in zooxanthellae division would not appear to be an overtly adverse effect although it may result in a shift in material transfer between the host and symbiont. Blooms of endosymbionts could draw more heavily on host resources and exacerbate hyperoxia. Zooxanthellae density is a convenient measurement for field biomonitoring and may show a different response with other toxicants.

The advantages of *A. elegantissima* as a biomonitoring organism lie largely with the attributes of the species, which are in good accordance with the biomonitoring species selection criteria outlined in the introduction. Based on the results of this study, several endpoints warrant further development for inclusion into biomonitoring programs. Endpoints were selected at several levels of biological organization: bioaccumulation, biochemical, physiological, behavioral, and community (endosymbiotic zooxanthellae). Growth inhibition, behavior, zooxanthellae division, and copper uptake may be easily measured from anemones kept in the field or laboratory. Many of these endpoints meet the endpoint criteria listed in the introduction, as well. The dose response sensitivity was not within the range of copper concentrations typically found at polluted sites, however, appropriate sensitivity may be found through other experimental designs or with other toxicants. The question remains how this species may be effectively utilized as a biomonitoring tool.

As was touched upon above, biomonitoring programs may assume a variety of forms depending on the goals and budget of the investigation. The monitoring program may be targeted to pollutant concentrations in biological tissues, or may examine effects at several levels of biological organization. Clearly the most efficient design generates the most useful information for the lowest level of effort.

While the response of one species is fundamentally not adequate to estimate ecosystem level pollutant impacts, the inclusion of a variety of species and endpoints into the examination will yield increasing accuracy. To this end, employment of toxicity tests utilizing *A. elegantissima* provides greater confidence in evaluations of pollutant impacts in the intertidal zone. Aside from this simple adjunction to the existing suite of toxicity tests, the species may provide a more efficient system for biomonitoring, in particular through *in-situ* deployments.

The U.S. EPA Marine/Estuarine Complex Effluent Toxicity Testing Program (U.S. EPA, 1989) approaches effluent biomonitoring utilizing a tiered approach, with the level of effort determined by the results of the tests in the preceding tier. Measurements of pollutant uptake, shell growth and scope for growth in *Mytilus edulis* deployed *in-situ* is used to complement laboratory toxicity testing of effluents and receiving waters. *In-situ* monitoring has the advantage of providing long-term site specific estimates of pollutant impacts. Variations in seasonal environmental factors and pollutant input rates are integrated over the deployment period. *Anthopleura elegantissima* may serve well as a replacement, or supplement, to the use of *Mytilus edulis*. Furthermore, *A. elegantissima* may be well suited to monitoring intertidal pollution through *in-situ* deployment as its tissues are continuously exposed, and clonal replicates may be utilized. Few, if any, pollution biomonitoring programs target the vast rocky intertidal zones of the world.

The goal of biomonitoring is to generate information within a conceptual framework that is useful for managing human impacts to ecosystems. Effective biomonitoring requires the knowledge of relationships between measurement endpoints and ecosystem level responses over long time periods. Developing economical and efficient biological effects measurement systems, including suitable *in-situ* biomonitoring species, facilitates the goal of effective ecosystem management.

CONCLUSIONS

1. The *Anthopleura elegantissima* 28 day lethality rangefinding test resulted in a copper sulfate EC_{50} estimate of 1350 µg/L.

2. Anemone growth, tentacle retraction behavior, feeding aversion behavior, and endosymbiotic zooxanthellae division rate,responses occurred at concentrations well below that of the copper sulfate EC_{50}.

3. Copper was bioaccumulated linearly with dose.

4. Zooxanthellae division was stimulated with increasing copper dose.

4. The existence of an overall treatment group dose response was confirmed by multivariate analysis of variance.

5. The biological attributes of *Anthopleura elegantissima* are largely in accordance with established criteria for selection of biomonitoring species.

6. Based on this initial evaluation of *Anthopleura elegantissima*, the species is amenable to toxicity testing and use in environmental pollution biomonitoring programs.

REFERENCES CITED

AOAC. 1980. Atomic absorbtion method for fish. In *Official Methods of Analysis of the Association of Analytical Chemists,* 13th ed. p399.

Anderson, D.M. and F.M.M Morel. 1978. Copper sensitivity of *Gonyaulax tamarensis. Limnology and Oceanography.* 23(2):283-295.

Aruoma, O.I., B. Halliwell, E. Gajewski and M. Dizdaroglu. 1991. Copper-ion-dependent damage to the bases in DNA in the presence of hydrogen peroxide. *Biochem. J.* 273:601-604.

Ates, R.M.L. 1991. Predation on cnidaria by vertebrates other than fishes. *Hydrobiologia.* 216/217:305-307.

Bay, S., R. Burgess and D. Nacci. 1993. Status and applications of echinoid (phylum echinodermata) toxicity test methods. In W.G. Landis, J.S. Hughes and M. Lewis, eds., *Environmental Toxicology and Risk Assessment, ASTM STP 1179.* American Society for Testing and Materials, Philadelphia, PA, pp. 281-302.

Beers, R.F. Jr. and I.W. Sizer. 1952. A spectrophotometric method for measuring the breakdown of hydrogen peroxide by catalase.*J. Biol. Chem.*195:133-140.

Black, R.E. and L. Bloom. 1984. Heat shock proteins in *Aurelia* (Cnidaria, Scyphozoa). *Jour. Exp. Zool.* 230:303-307.

Bodammer, J.E. 1979. Preliminary observations on the cytopathological effects of copper sulfate on the chemoreceptors of *Callinectes sapidus.* In W.B. Vernberg, A. Calabrese, F.P. Thurberg and F.J. Vernberg, eds., *Marine Pollution: Functional Responses.* Academic Press. New York, NY, pp. 223-237.

Borg, D.C. and K.M Schaich. 1984. Cytotoxicity from coupled redox cycling of autoxidizing xenobiotics and metals. *Isreal Journal of Chemistry.* 24:38-53.

Bowmer, T., R.G.V. Boelens, B.F. Keegan and J. O'Neill. 1986. The use of marine benthic key species in ecotoxicological testing: *Amphiura filiformis* (O.F. Müller) (Echinodermata: Ophiuroidea). *Aquatic Toxicology.* 8:93-109.

Boyle, E.A., and J.A. Edmond. 1975. Copper in surface waters south of new zealand. *Nature.* 253:270.

Boyle, E.A. 1979. Copper in natural waters. In J.O. Nriagu, ed., *Copper in the Environment, Part 1, Ecological Cycling.* John Wiley & Sons, New York, NY, pp. 77-88.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72:248-254.

Brand, L.E., W.G. Sunda and R.R.L. Guillard. 1986. Reduction of marine phytoplankton reproduction rates by copper and cadmium. *J. Exp. Mar. Biol. Ecol.* 96:225-250.

Brown, B.E. and S. Howard. 1985. Responses of coelenterates to trace metals: A field and laboratory evaluation. *Proc. 5th Int. Reef Congress.* 6:465-470.

Buckley, J.T., M. Roch, J.A. McCarter, C.A. Rendell and A.T. Matheson. 1982. Chronic exposure of coho salmon to sublethal concentrations of copper-I. Effect on growth, on accumulation and distribution of copper on copper tolerance. *Comp. Biochem. Physiol.* 72(1):15-19.

Bus, J.S. and J.E. Gibson. 1979. Lipid peroxidation and its role in toxicology. *Reviews in Biochemical Toxicology.* 1:125-149.

Cairns, J.C. Jr. 1986. The myth of the most sensitive species. *Bioscience,* 36(10):670-672.

Capuzzo, J.M. and J.J. Sasner. 1977. The effect of chromium on filtration rate and metabolic activity of *Mytilus edulis* L. and *Mya arenaria* L. In F.J. Vernberg, A. Calabrese, F.P. Thurberg and W.B. Vernberg, eds., *Physiological Responses of Marine Biota to Pollutants.* Academic Press, New York, NY. pp. 225-237.

Chan, P.C, O.G. Peller and L. Kesner. 1982. Copper(II)-catalyzed lipid peroxidation in liposomes and erythrocyte membranes. *Lipids.* 17(5):331-337.

Conklin, E.J. and R.N. Mariscal. 1976. Increase in nematocyst and spirocyst discharge in a sea anemone in response to mechanical stimulation. In G.O. Mackie, ed. *Coelenterate Ecology and Behavior.* Plenum Press, New York, NY. pp.

Døving, K.B. 1991. Assessment of animal behavior as a method to indicate environmental toxicity. *Comp. Biochem. Physiol.* 100C(1/2):247-252.

Dykens, J.A. 1984. Enzymic defenses against oxygen toxicity in marine cnidarians containing endosymbiotic algae. *Marine Biology Letters.* 5:291-301.

Dykens, J.A and J.M. Shick. 1982. Oxygen production by endosymbiotic algae controls superoxide dismutase activity in their animal host. *Nature.* 297:579-580.

Dykens, J.A., J.M. Shick, C. Benoit, G.R. Buettner and G.W. Winston. 1992. Oxygen radical production in the sea anemone *Anthopleura elegantissima* and its endosymbiotic algae. *J. Exp. Biol.* 168:219-241.

Ettinger, M.J. 1984. Copper metabolism and diseases of copper metabolism. In R. Lontie, ed., *Copper Proteins and Copper Enzymes.* CRC Press, Boca Raton, FL, pp. 175-229.

Federal Register, 1984. Development of water quality based permit limitations for toxic pollutants; national policy. 49(48), Friday, March, 9.

Fitt, W.K., R.L. Pardy and M.M. Littler. 1982. Photosynthesis, respiration, and contribution to community productivity of the symbiotic sea anemone *Anthopleura elegantissima* (Brandt, 1835). *J. Exp. Mar. Biol. Ecol.* 61:213-232.

Foote, C.S. 1976. Photosensitized oxidation and singlet oxygen: consequences in biological systems. In W.A. Pryor, ed., *Free Radicals in Biology.* Academic Press, New York, NY, pp. 85-133.

Foote, C.S. 1980. Photooxidation and toxicity. In R.S. Bhatnagar, ed., *Molecular Basis of Environmental Toxicology.* Ann Arbor Science, Ann Arbor, MI, pp. 37-50.

Ford, C.E. 1964. Reproduction in the aggregating sea anemone *Anthopleura elegantissima. Pacific Science.* 18:138-145.

Francis, L. 1976. Social organization within clones of the sea anemone *Anthopleura elegantissima. Biol. Bull.* 150:361-376.

Francis, L. 1979. Contrast between solitary and clonal lifestyles in the sea anemone *Anthopleura elegantissima. Amer. Zool.* 19:669-681.

Fredericks, C.A. 1976. Oxygen as a limiting factor in phototaxis and in intraclonal spacing of the sea anemone *Anthopleura elegantissima. Marine Biology.* 38:25-28.

Freedman, J.H., M.R. Ciriolo and J. Peisach. The role of glutathione in copper metabolism and toxicity. *Journal of Biological Chemistry.* 264(10):5598-5605.

Glynn, P.W. 1984. Widespread coral mortality and the 1982-1983 El Niño warming event. *Envir. Conserv.* 10:149-154.

Goldberg, E.D., V.T. Bowen, J.W. Farrington, G. Harvey, J.H. Martin, P.L. Parker, R.W. Risebrough, W. Robertson, E. Schneider and E. Gamble. 1978. The mussel watch. *Environ. Conserv.* 5:101-125.

Hand, C. 1955. The sea anemones of Central California. Part II. The endomyarian and mesomyarian anemones. *Wasmann J. Biol.* 13:37-99.

Hansen, D.J. 1984. Utility of toxicity tests to measure effects of substances on marine organisms. In H.H. White, ed.,*Concepts in Marine Pollution Measurements.* Maryland Sea Grant College, University of Maryland, College Park, MD, pp. 33-56.

Harland, A.D. and N.R. Nganro. 1990. Copper uptake by the sea anemone *Anemonia viridis* and the role of zooxanthellae in metal regulation. *Marine Biology*.104:297-301.

Harland, A. D., G.W. Bryan and B.E. Brown. 1990, Zinc and cadmium absorption in the symbiotic anemone *Anemonia viridis* and the non-symbiotic anemone *Actinia Equina*. *J. Mar Biol. Ass. U. K.* 70:789-802.

Houvenaghel, G.T. 1984. The hydroid *Hydractinia echinata* as a suitable bioassay species for low reacting compounds. In G. Persoone, E. Jaspers, and C. Claus, eds., *Ecological Testing for the Marine Environment Vol 2.* State Univ. Ghent and Inst. Mar. Scient. Res., Bredene, Belgium. pp. 67-83.

Jackim, E. 1974. Enzyme responses to metals in fish. In F.J. Vernberg and W.B. Vernberg, eds., *Pollution and Physiology of Marine Organisms.* Academic Press, New York, NY, pp. 59-65.

Jennison, B.L. 1979. Annual fluctuations of the lipid levels in the sea anemone *Anthopleura elegantissima* (Brandt, 1835). *J. Exp. Mar. Biol. Ecol.* 39:211-221.

Jewell, C.S.E. and G.W. Winston. 1989. Oxyradical production by hepatopancreas microsomes from the red swamp crayfish, *Procambarus clarkii. Aquat. Toxicol.* 14:27-46.

Johnson, E. M. 1983. In M.S. Christian, W.M. Galbraith, P. Voytek, and M.A. Mehlman, eds., *Advances in Modern Environmental Toxicology III*, Princeton Science Publications, Princeton, NJ, pp. 77-91.

Jop, K. M., A.M. Askew, D.J. Texeira and J. MacGregor. 1993. Quality control in aquatic toxicity testing programs: evaluation of copper and hexavalent chromium as reference toxicants. In W.G. Landis, J.S. Hughes, and Ml. Lewis, eds., *Environmental Toxicology and Risk Assessment, ASTM STP 1179*,. American Society for Testing and Materials, Philadelphia, PA, pp. 397-404.

Karbe, L. 1972. Marine hydroiden als testorganismen zur prufung der toxizitat von abwasserstoffen. Die wirkung von schwermetallen auf kolonien von *Eirnen viridula. Marine Biology.* 12:316-328.

Karbe, L., T. Borchardt, R. Dannenberg and E. Meyer. 1984. Ten years of experience using marine and freshwater hydroid bioassays. In G. Persoone, E. Jaspers and C.Claus, eds., *Ecological Testing for the Marine Environment, Vol. 2.* State Univ. Ghent and INst. Mar. Scient. Res., Bredene, Belgium. pp.99-129.

Lesser, M.P. 1989. Photobiology of natural populations of zooxanthellae from the sea anemone *Aiptasia pallida*: assessment of the hosts role in protection against ultraviolet radiation. *Cytometry.* 10:653-658.

Lesser, M.P. and J.M. Shick. 1989. Photoadaptation and defenses against oxygen toxicity in zooxanthellae from natural populations of symbiotic cnidarians. *J. Exp. Mar. Biol. Ecol.*. 134:129-141.

Lindstedt, K.J. 1971. Chemical control of feeding behavior. *Comp. Biochem. Physiol.* 39A:553-581.

Littler, M.M. 1980. Southern california rocky intertidal ecosystems: methods, community structure and variability. In J.H. Price, D.E.G. Irvine and W.F. Farnham, eds.,*The ShoreEnvironment, Vol. 2, Ecosystems*, Academic Press, New York, NY, pp. 565-608.

Livingstone, D.R., P. Garcia Martinez, X. Michel, J.F. Narbonne, S. O'Hara, D. Ribera and G.W. Winston. 1990. Oxyradical production as a pollution mediated mechanism of toxicity in the common mussel *Mytilus edulis* L., and other molluscs. *Functional Ecology.* 4:415-424.

Macfarland, F.K. and G. Muller-Parker. 1993. Photosynthesis and retention of zooxanthellae and zoochlorellae within the aeolid nudibranch *Aeolida papillosa. Biol. Bull.* 184:223-229.

Mariscal, R.N. 1974. Nematocysts. In L. Muscatine and H.M. Lenhoff, eds., *Coelenterate Biolology: Reviews and New Perspectives.* Academic Press, New York, NY. pp. 129-178.

Martin, R.B. 1986. Bioinorganic chemistry of metal ion toxicity. In H. Sigel, ed., *Metal Ions in Biological Systems, Vol. 20: Concepts on Metal Toxicity.* Marcel Dekker, Inc., New York, NY, pp. 21-65.

Martin, M. and R. Severeid. 1984. Mussel watch monitoring for the assessment of trace toxic constituents in California marine waters. In H.H. White. ed.,*Concepts in Marine Pollution Measurements.* Maryland Sea Grant, University of Maryland, College Park, MD, pp. 291-323.

Matthews, G. and J. Hearne. 1991. Clustering without a metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 13(2):175-184.

McKnight, D.M. and F.M.M Morel. 1979. Release of weak and strong copper complexing agents by algae. *Limnol. Oceanogr.* 24(5):823-827.

Minkel, D.T., K. Poulsen, S. Wielgus, C.F. Shaw, III and D.H. Petering. 1980. On the sensitivity of metallothioneins to oxidation during isolation. *Biochem. J.*, 191:475-485.

Moore, M.N. and A.R.D. Stebbing. 1976. The Quantitative cytochemical effects of three metal ions on a lysosomal hydrolase of a hydroid. *Journal Mar. Biol. Ass. U. K.* 56:995-1005.

Moore, M.N. 1980. Cytochemical determination of cellular responses to environmental stressors in marine organisms. *Rapp. P.-v. Réun. Cons. int.Explor. Mer.* 179:7-18.

Muscatine, L. 1961. Symbiosis in marine and freshwater coelenterates. In H.M. Lenhoff and W.F. Loomis, eds., *The Biology of Hydra.* University of Miami Press, Coral Gables, FL, pp. 255-268.

Nath, K.A. and A.K. Salahudeen. 1993. Autooxidation of cysteine generates hydrogen peroxide: cytotoxicity and attenuation by pyruvate. *Am. J. Physiol.* 264:F306-F314.

Okazaki, R.K. 1976. Copper toxicity in the pacific oyster *Crassostrea gigas. Bull. Environ. Contam. Toxicol.* 16:658-664.

Paine, R.T. 1966. Food web complexity and species diversity. *The American Naturalist.* 100(910):65-75.

Pearse, V.B. 1974. Modification of sea anemone behavior by symbiotic zooxanthellae: expansion and contraction. *Biol. Bull.* 147:641-651.

Peddicord, R.K. 1984. What is the meaning of bioaccumulation as a measure of marine pollution effects?. In H.H. White. ed.,*Concepts in Marine Pollution Measurements.* Maryland Sea Grant. University of Maryland, College Park, MD, pp. 291-323.

Perkin Elmer, 1982, Analysis of fish and seafood. In *Analytical Methods for Atomic Absorption Spectrophotometry.* Perkin Elmer Corporation, Norwalk, CT.

Petering, D.H. and W.E. Antholine. 1988. Copper toxicity: Speciation and reactions of copper in biological systems. *Reviews in Biochemical Toxicology.* 9:225-270.

Phillips, D.J.H.1980. *Quantitative Aquatic Biological Indicators-Their Use to Monitor Trace Metal and Organochlorine Pollution,* Applied Science Publ. Ltd., London, p.488.

Powell, M.I. and K.N. White. 1989. Influence of the heavy metals copper and cadmium on the behaviour of *Semibalanus balanoides* and *Balanus crenatus. Mar. Behav. Physiol.* 14:115-127.

Radi, A.A.R. and B. Matkovics. 1988. Effects of metal ions on the antoxidant enzyme activities, protein contents and lipid peroxidation of carp tissues. *Comp. Biochem. Physiol.* 90C(1):69-72.

Redpath, K.J. and J. Davenport. 1988. The effect of copper, zinc and cadmium on the pumping rate of *Mytilus edulis* L. *Aquatic Toxicology.* 13:217-226.

Regoli, F. 1992. Lysosomal responses as a sensitive stress index in biomonitoring heavy metal pollution. *Mar. Ecol. Prog. Ser.* 84:63-69.

Reish, D.J., J.M. Martin, F.M. Piltz and J.Q. Word. 1976. The effect of heavy metals on laboratory populations of two polychaetes with comparisons to the water quality conditions and standards in southern california marine waters. *Water Res.* 10:299-302.

Rijstenbil, J.W. and J.A. Wijnholds. 1991. Copper toxicity and adaptation in the marine diatom *Dictylum brightwelli. Comp. Biochem. Physiol.* 100C(1/2):147-150.

Samuni, A. M. Chevion and G. Czapski. 1981. Unusual copper-induced sensitization of the biological damage due to superoxide radicals. *The Journal of Biological Chemistry.* 256(24):12632-12635.

Scott, D. and C. Major. 1972. The effect of copper II on survival, respiration, and heart rate in the common blue mussel *Mytilus edulis. Biol. Bull.* 143:679-688.

Sebens, K.P. 1979. The energetics of asexual reproduction and colony formation in benthic marine invertebrates. *Am. Zool.*, 19:683-697.

Sebens, K.P. 1980. The control of body size and longitudinal fission in the sea anemone *Anthopleura elegantissima. Biol. Bull.* 158:370-382.

Sebens, K.P. 1981a. Reproductive ecology of the intertidal sea anemones *Anthopleura elegantissima* (Brandt) and *A. xanthogrammica* (Brandt): Body size, habitat, and sexual reproduction. *J. Exp. Mar. Ecol.* 54:225-250.

Sebens, K.P. 1981b. The allometry of feeding, energetics, and body size in three sea anemone species. *Biol. Bull.* 161:152-171.

Sebens, K.P. 1982. Recruitment and habitat selection in the intertidal sea anemones, *Anthopleura elegantissima* (Brandt) and *A. xanthogrammica* (Brandt). *J. Exp. Mar. Biol. Ecol.* 59:103-124.

Sebens, K.P. 1983. Morphological variability during longitudinal fission of the intertidal sea anemone *Anthopleura elegantissima* (Brandt). *Pacific Science.* 37(2):121-132.

Shelton, G.A.B.1982. Anthozoa. In G.A.B. Shelton, ed., *Electrical Conduction and Behavior in "Simple" Invertebrates.* Oxford Univ. Press, New York, NY, pp. 203-242.

Shick, J.M. 1991. *A Functional Biology of Sea Anemones.* Chapman & Hall, New York, NY.

Shick, J.M., Hoffman, R.J., and Lamb, A.N. 1979. Asexual reproduction, population structure, and genotype-environment interactions in sea anemones. *Amer. Zool.* 19:699-713.

Shick, J.M. and J.A. Dykens. 1984. Photobiology of the symbiotic sea anemone *Anthopleura elegantissima*: Photosynthesis, respiration, and behavior under intertidal conditions. *Biol. Bull.* 166:608-619.

Shick, J.M. and J.A. Dykens. 1985. Oxygen detoxification in algal-invertebrate symbioses from the Great Barrier Reef. *Oecologia.* 66:33-41.

Shinar, E., T. Navok and M. Chevion. 1983. The analogous mechanisms of enzymatic inactivation induced by ascorbate and superoxide in the presence of copper. *The Journal of Biological Chemistry.* 258(24):14778-14783.

Singh, S.M. and P.M. Sivalingham. 1982. *In vitro* study on the interactive effects of heavy metals on catalase activity of *Sarotherodon mossambicus* (Peters). *J. Fish. Biol.* 20:683-688.

SPSS, Inc. 1992. *SPSS Release 4.0,* SPSS, Inc., Chicago, IL.

Stebbing, A.R.D. 1976. The effects of low metal levels on a clonal hydroid. *J. Mar. Biol. Ass. U.K.* 56:977-994.

Stebbing, A.R.D. 1979. An experimental approach to the determinants of biological water quality. *Phil. Trans. R. Soc. Lond.* B. 286:465-481.

Stebbing, A.R.D. 1981. The effects of reduced salinity on colonial growth and membership in a hydroid. *J. Exp. Mar. Biol. Ecol.* 55:233-241.

Stebbing, A.R.D., J.J. Cleary, M. Brinsley, C. Goodchild and V. Santiago-Fandiño. 1983. Responses of a hydroid to surface water samples from the river Tamar and Plymouth Sound in relation to metal concentrations. *J. Mar. Biol. Ass. U.K.* 63:695-711.

Stebbing, A.R.D. and B.E. Brown. 1984. Marine ecotoxicological tests with coelenterates. In G. Persoone, E. Jaspers and C.Claus, eds., *Ecological Testing for the Marine Environment, Vol. 1* . State Univ. Ghent and Inst. Mar. Scient. Res., Bredene, Belgium, pp.307-339.

Sternlieb, I. and S. Goldfischer. 1976. Heavy metals and lysosomes. In, J.T. Dingle and R.T. Dean, eds., *Lysosomes in Biology and Pathology, Vol. 5.* North Holland, Amsterdam, pp. 185-200.

Sunda, W. and R.R.L. Guillard. 1976. The relationship between cupric ion activity and the toxicity of copper to phytoplankton. *Journal of Marine Research.* 34(4):511-529.

Tapley, D.W. 1989. Photoinactivation of catalase but not superoxide dismutase in the symbiotic sea anemone *Aiptasia pallida*. *Amer. Zool.* 29:53A.

Thursby, G. B., B.S. Anderson, G.E. Walsh and R.L. Steele. 1993. A review of the current status of marine algal toxicity testing in the united states. In W.G. Landis, J.S. Hughes and M. Lewis, eds.,*Environmental Toxicology and Risk Assessment, ASTM STP 1179.* American Society for Testing and Materials, Philadelphia, PA, pp. 362-377.

Toxstat, 1990. *Toxstat Release 3.2*, D.D. Gulley, A.M. Boelter and H.L. Bergman, University of Wyoming, Laramie, WY.

U.S. EPA. 1989. *Biomonitoring for the Control of Toxicity in Effluent Discharges to the Marine Environment.* EPA/625/8-89/015. U.S. EPA Center for Environmental Research Information-Office of Research and Development, Cincinnati, OH.

Van der Schalie, W.H. 1986. Can biological monitoring early warning systems be useful in detecting toxic materials in water? In T.M. Poston and R. Purdy, eds.,*Environmental Toxicology and Risk Assessment, ASTM STP 921,* American Society for Testing and Materials, Philadelphia, PA, pp. 362-377.

Viarengo, A., G. Zanicchi, M.N. Moore and M. Orunesu. 1981. Accumulation and detoxification of copper by the mussel *Mytilus galloprovincialis* Lam.: a study of the subcellular distribution in the digestive gland cells. *Aquatic Toxicology.* 1:147-157.

Viarengo, A., L. Canesi, M. Pertica, G. Poli, M.N. Moore and M. Orunesu. 1990. Heavy metal effects on lipid peroxidation in the tissues of *Mytilus galloprovincialis* Lam. *Comp. Biochem. Physiol.* 97C:37-42.

Webber, C.I. 1991. *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms, 4th Edition.* United States Environmental Protection Agency, Office of Research and Development, Washington D.C.

White, H.H. 1984. Mussel madness: Use and misuse of biological monitors of marine pollution. In H.H. White, ed.,*Concepts in Marine Pollution Measurements.* University of Maryland, College Park, MD, pp. 325-337.

Widdows, J. and P. Donkin. 1991. Role of physiological energetics in ecotoxicology. *Comp. Biochem. Physiol.* 100C(1/2):69-75.

Wilkerson, F. P., G. Muller-Parker, L. Muscatine. 1983. Temporal patterns of cell division in natural populations of endosymbiotic algae. *Limnol. Oceanogr.* 28(3):1009-1014.

Winston, G.W. 1991. Oxidants and antioxidants in aquatic animals. *Comp. Biochem. Physiol.* 100C(1/2):173-176.

Zachariassen, K.E., T. Aunaas, J.F. Børseth, S. Einarson, T. Nordtug, A. Olson and G. Skjervø. 1991. Physiological parameters in ecotoxicology. *Comp. Biochem. Physiol.* 100C(1/2):77-79.

Zamer, W.E. and J.M. Shick. 1987. Physiological energetics of the intertidal sea anemone *Anthopleura elegantissima*: II. energy balance. *Marine Biology.* 93:481-491.

Zamer, W.E. and J.M. Shick. 1989. Physiological energetics of the intertidal sea anemone *Anthopleura elegantissima*, III. Biochemical composition of body tissues, substrate specific absorption, and carbon and nitrogen budgets. *Oecologia.* 79(1):117-127.