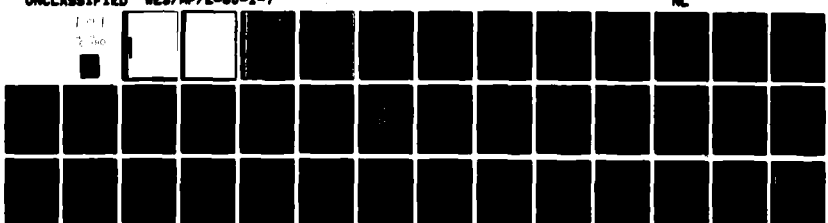
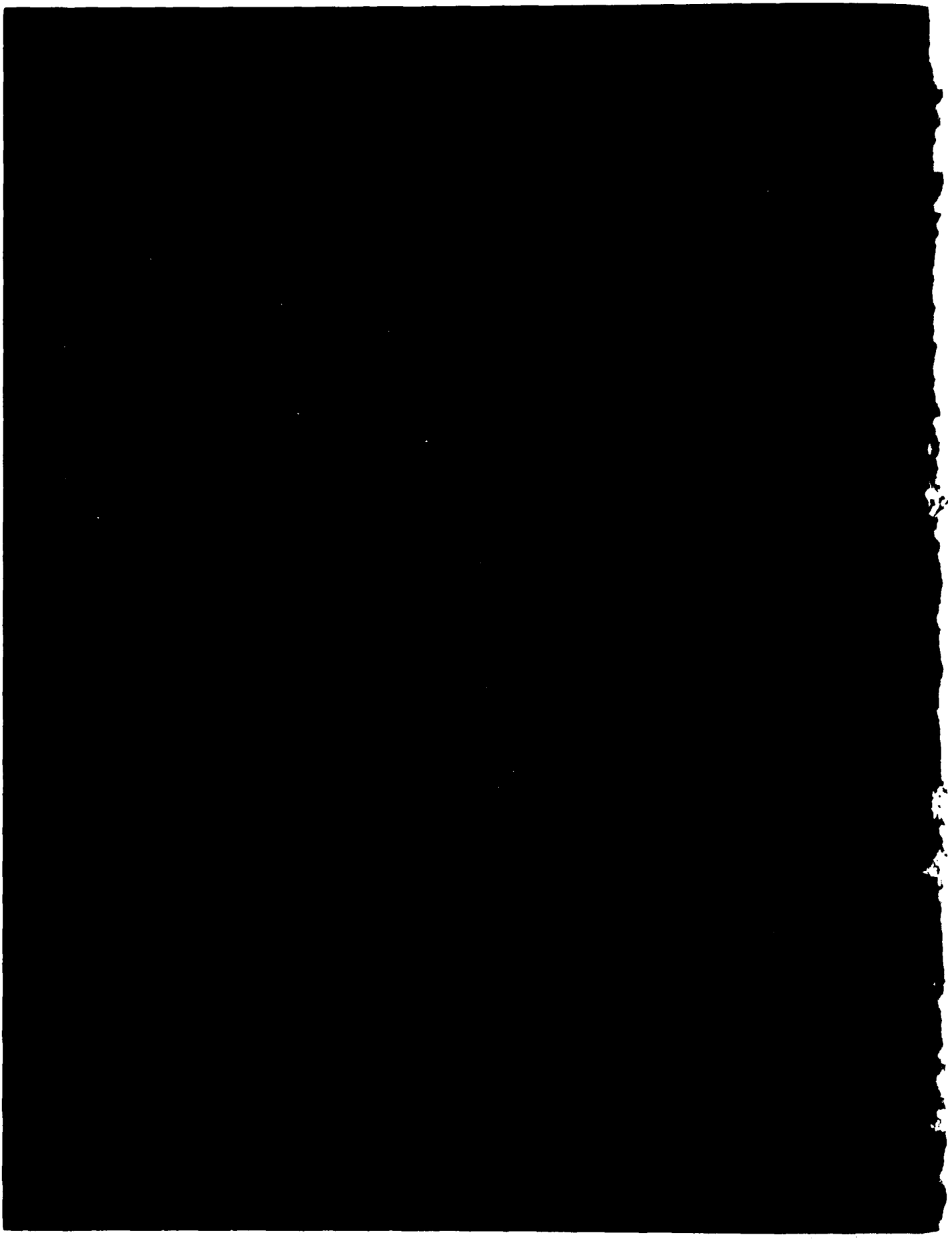
AD-A116 990 ARMY ENGINEER WATERWAYS EXPERIMENT STATION VICKSBURG--ETC F/G 13/2
AQUATIC HABITAT STUDIES ON THE LOWER MISSISSIPPI RIVER, RIVER M--ETC(U)
FEB 82 M P FARRELL, A D MAGOUN, K DANIELS

UNCLASSIFIED WES/MP/E-80-1-7 NL



END
DATE FILMED
4-82
DTIC



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Miscellaneous Paper E-80-1	2. GOVT ACCESSION NO. AD-A116 990
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): AQUATIC HABITAT STUDIES ON THE LOWER MISSISSIPPI RIVER, RIVER MILE 480 TO 530; Report 7, MANAGEMENT OF ECOLOGICAL DATA IN LARGE RIVER ECOSYSTEMS	5. TYPE OF REPORT & PERIOD COVERED: Report 7 of a series
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Michael P. Farrell, A. Dale Magoun, Karen Daniels, C. H. Pennington, and R. H. Strand	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: U. S. Army Engineer Waterways Experiment Station, Environmental Laboratory, P. O. Box 631, Vicksburg, Miss. 39180	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: EWQOS Work Unit VIIB
11. CONTROLLING OFFICE NAME AND ADDRESS: Office, Chief of Engineers, U. S. Army, Washington, D. C. 20314	12. REPORT DATE: February 1982
	13. NUMBER OF PAGES: 35
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Available from National Technical Information Service, 5285 Port Royal Road, Springfield, Va. 22151.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Aquatic ecology
Computer programming management
Environmental engineering
Mississippi River
Water quality

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

To expedite the research data management system (RDMS) required for complex and large-scale ecological field studies being done at the U. S. Army Engineer Waterways Experiment Station (WES), a graphical display system was developed. The Statistical Analysis System (SAS) provides the framework for making open-ended RDMS practical and cost-effective. PROC VIVIPLOT is the first of a series of SAS procedures that will produce copy-ready figures with

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued).

some independence as to choice of plotter. The system has been developed by the McDonnell Douglas Automation Company (MCAUTO) at the request of personnel at WES. The use of PROC VIVIPLOT in the MCAUTO environment, including plotting characteristics and cost, are discussed.

The SAS procedure VIVIPLOT produces an Intermediate Plot File (IPF)-(MCAUTO) from IPF system routines that generate a data file which is post-processed to produce a file to drive Calcomp (748, 663, 960), Xynetics and Houston pen and ink plots, Gould 4800 and Varian electrostatic plots, and displays Computek and Tektronix storage tubes.

Since the capabilities and syntax of PROC VIVIPLOT are nearly identical to PROC PLOT, line plots from PLOT can be easily converted to copy-ready figures using VIVIPLOT. Several enhancements to the capabilities of PROC PLOT have been made in PROC VIVIPLOT and include multiple vertical and horizontal axis labeling, reserved area position and labeling, and connected lines from data vectors in SAS data sets.

In studies such as the Environmental and Water Quality Operational Studies (EWQOS) that utilize multiple data bases composed of hierarchical file structures, there is a high probability that errors may be perpetuated into summary reports unless some form of quality assurance is integrated into the research data base management program. In studies that substitute numeric codes for character variable values, this problem of error propagation is even more acute. This report addresses the problem of error propagation in those studies employing a coding scheme to represent longer alphanumeric values.

Several approaches are available that minimize errors in coding variables. Numeric codes, "smart codes," with embedded information allocated to positions within the value codes are widely used but unacceptable for variables with many values and/or many levels of classification. "Nonsense" codes, or codes without embedded information, however, efficiently circumvent the problems associated with smart codes. Using nonsense codes, alphanumeric variable values are assigned a sequential numeric code as new values are encountered in the data base, irrespective of the position of the value in the classification scheme for that variable. With the use of nonsense codes, the management approach is open-ended and does not require a knowledge of the number of potential classification levels for the variables. In addition, experience with several large environmental data bases indicates that coding errors appear to be less frequent using nonsense codes than in those studies in which a smart code approach was used.

This is Report 7 of the series "Aquatic Habitat Studies on the Lower Mississippi River, River Mile 480 to 530." A complete listing of the reports is as follows:

Report 1: Introduction
Report 2: Aquatic Habitat Mapping
Report 3: Benthic Macroinvertebrate Studies--Pilot Report
Report 4: Diel Periodicity of Benthic Macroinvertebrate Drift
Report 5: Fish Studies--Pilot Report
Report 6: Larval Fish Studies--Pilot Report
Report 7: Management of Ecological Data in Large River Ecosystems
Report 8: Summary

PREFACE

The study reported herein was conducted as part of the Environmental and Water Quality Operational Studies (EWQOS), Work Unit VIIB, Waterway Field Studies conducted by the Environmental Systems Division (ESD), Environmental Laboratory (EL), U. S. Army Engineer Waterways Experiment Station (WES). The EWQOS Program is sponsored by the Office, Chief of Engineers, and is assigned to WES under the purview of EL. This is Report 7 of the series "Aquatic Habitat Studies on the Lower Mississippi River, River Mile 480 to 530." The study was undertaken from April to October 1978.

This report summarizes the initial development of a Environmetrics Program (EP) for the Long Term Field Studies of EWQOS.

The contribution of the personnel of the Waterway Habitat and Monitoring Group (WHMG) for their patience during the genesis of this data base management program is acknowledged. Support for the senior author during the development of the management system was made possible by an IPA appointment with the U. S. Army Corps of Engineers from Miami University, *Oxford, Ohio.*

The study was under the supervision of Dr. Walter B. Gallaher, former Chief, WHMG; Dr. Thomas D. Wright, Chief, WHMG; Mr. Bob Benn, Chief, ESD; Dr. Jerry Mahloch, Program Manager, EWQOS; and Dr. John Harrison, Chief, EL.

COL John L. Cannon, CE, and COL Nelson P. Conover, CE, were Commanders and Directors of WES during the conduct of this study and the preparation of this report. Mr. Fred R. Brown was Technical Director.

This report should be cited as follows:

> Farrell, M. P., Magoun, A. D., Daniels, K., Pennington, C. H., and Strand, R. H. 1982. "Aquatic Habitat Studies on the Lower Mississippi River, River Mile 480 to 530; Report 7, Management of Ecological Data in Large River Ecosystems," Miscellaneous Paper E-80-1, U. S. Army Engineer Waterways Experiment Station, CE, Vicksburg, Miss.

CONTENTS

AQUATIC HABITAT STUDIES ON THE LOWER MISSISSIPPI RIVER, RIVER MILE 480 TO 530

MANAGEMENT OF ECOLOGICAL DATA IN LARGE RIVER ECOSYSTEMS

PART I: PROBLEMS IN APPLYING ENVIRONMETRICS TO THE ENVIRONMENTAL SCIENCES

Introduction

The general problem

1. The foundation of sound research data base (RDB) management is detailed planning (Martin 1976). One who manages such a data base is usually barraged by the apparent need for flowcharts, PERT diagrams, cross-reference libraries, directional dictionaries, and a host of other "aids" designed to increase efficiency and/or ensure a final product that will accomplish the goals of the project. Plans for recovering data sets, sorting strategies, merging and updating capabilities, and accomplishing intercomputer exchanges must be planned well in advance with few allowances for "lurking variables" (Box, Hunter, and Hunter 1979). The research data manager must also give detailed breakdowns on development time, personnel and computer costs, and the lead time necessary to develop the application programs even though the project may be several years away with the possibility of personnel turnover and/or hardware changes.

2. Informal polls among research data managers indicate that an increasing number of research data management systems (RDMS) are not being developed as outlined in many of the leading references in this field. For example, flowcharting, one of the backbones of the industry, has been shown to be an academic exercise with little application or help to real world complex data base problems. PERT diagrams are very informative after the final report is written. Dictionaries, fixed sorting strategies, cumbersome merging capabilities, data set recovery

problems, etc. are no longer problem areas due to exceptional hardware developments.

3. Furthermore, most research data managers find it difficult, if not impossible, to cost-justify more than a cursory research data management plan. Time estimates for development are usually too long and must be reduced. In addition, cost estimates may be very inaccurate when many research programs cannot identify all the variables or data base formats that will ultimately be necessary for analysis and report generation.

4. Even though the previously mentioned problems exist in research data management applications, a discernible naivety is associated with many of the currently published reports dealing with RDMS. One of the principle reasons for this discrepancy may be the different approaches in research data management found among applied contract researchers with strict budgetary constraints and university based research programs with their more liberal approach to computer-related costs. Although free computer time at the universities is becoming very rare, there still exists such a price differential as to perhaps subliminally encourage many of the data management strategies popularized in reference material.

5. Another area of concern in research data management is the problem demonstrated at its worst by the apparent trend among research data managers trained as programmers to view most projects as unique with unique solutions. The list of in-house developed data base management programs written in FORTRAN and/or COBOL specifically for a project must be horrendous. The experience of the authors of this report could supply a long list of these data base management structures so specific as to preclude any general use and, often, only meeting a small proportion of the needs of the RDB manager because of cost overruns, programming problems, or changes in the project's emphasis. On the other hand, this tendency to "reinvent the wheel" does not seem as popular among research data managers who have been trained in areas other than programming and who are aware of the new application programs currently available on lease or purchase options.

The environmental sciences and research data base management

6. Unlike many of the other sciences, the environmental sciences have most of the problems identified previously, with project goals often defined more appropriately after the study becomes operational. Many times at the startup of a project, variable selection and research data formats are often tentative because of the unknown biological complexity that may be encountered. Potential ways to summarize the data base are usually more numerous than money permits. Lead time for development of even a simple RDB structure is usually nonexistent. Research data managers frequently become involved with a project only shortly before data collection with the subsequent need for immediate data summarization so that the project may be modified before the next scheduled sampling period. In such an atmosphere, where the research data managers face a project that will provide answers by an iterative process, and where there will be major changes in the data base content and structure, the manager cannot hope to spend many days planning the specifics of the RDMS and only infrequently can the cost of development of such an RDMS be justified.

7. Faced with the uncertainties of managing an environmental data base, a research data manager is expected to provide a project with a skeleton of a data management system that can add broad ranges of new variables, reformat existing variables, perform analyses that are not anticipated, provide computer-generated tables and copy-ready figures that will be formatted at the conclusion of the study, and, in general, provide immediate answers via a time-sharing system, but provide the capabilities of reducing the cost of large complex analyses via batch operations. Superimposed on the above list, the system development must not demand excessively large and complex programming tasks. The system must be cost-effective with costs ranging from 5-10 percent of the total project funds--the lower the percentage, the better. Futhermore, the need for a system analyst to manage and construct the data base is by project definition "counterproductive." The requirement of having a research data manager is often considered an "unnecessary burden" to the

project and may jeopardize the project's financial capability to measure other "important variables" or eliminate some other aspect of the project. If the project group has a statistician, it is usually that person who is nominated to manage the data base.

8. Because of the complexity of the Mississippi River project and the fact that little information was available prior to the extensive field studies planned by the Corps, a six-month pilot study was initiated to help refine the general plan of study and to provide baseline data that was needed to evaluate variable selection and potential experimental designs. In relationship to the research data management aspects of the project, the situation faced by the project management team was quite familiar and reflected most of the difficulties of managing research data bases associated with changing extremely complex environmental studies as described previously.

The System Selection

9. Because of the research status of the U. S. Army Engineer Waterways Experiment Station (WES), the usual decision as to the choice of appropriate software based on hardware availability did not have to be entertained. WES has the ability to utilize in-house computing capabilities and/or purchase computer time from commercial vendors holding General Services Administration contracts. Therefore, the selection of an appropriate software system was of primary importance and was unencumbered by the usual hardware considerations.

10. To accomplish the short-term objectives of the pilot study and maintain the proper perspective of a long-term commitment to the overall study, the projected software system had to meet the following five major system selection criteria: (a) vendor support of the system's software including programming applications, analysis programs and help in troubleshooting user applications; (b) not only provide research data management capabilities that are easily programmed (user-oriented), flexible, and hierarchical in that canned instructions exist (e.g., sorting, merging, updating), but also user-programmed instructions (e.g., input, output, quality control checking); (c) provide a basic

complement of statistical analysis routines (i.e., means, standard deviation, analysis of variance, regression, etc.), plotting and charting capabilities, and more advanced programs that may be available in the system, programmable within the system, or available in other packages that interface with the parent system; (d) provide a common syntax for batch and time-shared operation; (e) be cost-effective, not only in terms of computer costs (e.g., core, central processing unit (CPU), input/output) but for the personnel time needed for implementation and maintenance.

11. After a review of project needs, an in-house development of application programs (i.e., data handling and analysis) was not judged a cost-effective alternative for this project. Considering development time estimates, personnel commitments, and costs, especially for the changing status of the pilot study, an in-house effort would immediately exceed one of the prime goals of keeping the RDMS at 5-10 percent of the project's funded level.

12. The Statistical Analysis System (SAS) was selected for use in this study because it adequately addressed each of the potential problem areas associated with both the pilot study and the long-term study. It also met favorably with the system selection criteria outlined previously.

Planning Versus Open-Ended Management

13. All data management structures must be planned. What is perhaps not clear is the amount and direction of planning necessary after an advanced analysis package has been selected. Detailed planning of research data bases appears to be inversely proportional to the degree to which the selected package meets the system selection criteria outlined previously. If adherence to the criteria is high, then planning the RDMS can be minimal with sessions devoted to determining output formats and requirements and any specialized analysis programs needed but not contained in the package. On the other hand, when a selected package has major omissions in relationship to the system selection criteria,

when adherence is low, planning time is usually increased with the emphasis on more of the basic problems of research data management such as variable input formats, internal file construction, sorting, merging, updating, etc. Therefore, adherence to the system selection criteria permits the research data manager to be more involved with the end-product requirements of the study, such as copy-ready graphical displays, computer-generated tables, and quality control assurances. In turn, the scientist involved with reporting the findings of the study benefits from this new end-product orientation of the research data manager. Now the scientists can become more involved with interpreting results of the study, as opposed to editorial requirements mandated without an efficaciously designed system. Furthermore, decisions that were previously based on inflexible computer program requirements can be modified so that the emphasis is placed on the needs of the scientists. As a result, efficiency is gained in the field operations where the majority of cost is usually involved without additional cost to the data management program.

14. The term "open-ended research data base management" (OE/RDBM) was coined to describe this philosophy of data management that supports a minimal planning effort and in which the emphasis is placed on computer-related needs at the completion of the study. Obviously, an OE/RDBM strategy will not work for all types of programs, nor for all levels of experience in executing the selected package. In addition, OE/RDBM may be precluded as a working system due to the emphasis of the program and/or the degree to which the selected package meets the system selection criteria. It appears that SAS is very amenable to OE/RDBM and that it reduces many of the potential hazards that could be encountered with an open-ended management system. The key element in using SAS for OE/RDBM, it was found, is a sound understanding of the structure of a SAS data set and the data set formats needed by a procedure to produce the desired results. Knowledge of the capabilities of SAS and of the lack of certain capabilities is also instrumental in developing an OE/RDBM system.

The Cost Factors

15. While OE/RDBM using SAS is an appealing alternative to research data management based on philosophical grounds, the cost differential of a OE/RDBM system may be even more attractive. The difficulty of assigning dollars to inflation factors, time estimates, computer equivalents, etc. is recognized, but even a simple, basic cost analysis of research data management alternatives provides some striking arguments for alternative approaches.

16. Table 1 shows the comparative cost analysis of a system using the OE/RDBM approach and SAS with a system developed within an organization at the program language level (FORTRAN). Both systems have been evaluated using commercial and private computer cost estimates. The manpower cost estimates are based on a skill level of a GS-12 rating including overhead burdens. All values are estimated on the basis of experience and reflect a detailed knowledge of the specific project's requirements, which are not presented in this paper. There are many other costs that could be included in the analysis (e.g., maintenance, storage, off-line and on-line devices, terminal leases, communication costs, programming support) but have been omitted because of the variable nature of these costs and in an effort to maintain simplicity.

17. The OE/RDBM approach and SAS can result in substantial savings to the project on the order of 50-60 percent, irrespective of the computer affiliation and with a minimal dollar outlay for an OE/RDBM system on a private computer (Table 1). As might be expected, analysis costs are approximately the same for both systems. The major difference between the two systems is the additional cost of manpower and computer time in the development of an in-house system--approximately 85 percent of the total research data management costs of an in-house system and only about 60 percent of the cost of an OE/RDBM system utilizing the capabilities of SAS. It should also be pointed out that the two systems would not be equal in terms of data base management or analysis capabilities. In this example, the OE/RDBM system using SAS would provide about 50 percent more capabilities in both management and analyses.

18. While these figures are tentative and reflect individual biases and experiences, there appears to be potential financial savings substantial enough to warrant detailed cost analysis at the local project level. For programs that have to be cost-effective, the OE/RDBM approach in conjunction with SAS will be difficult to ignore.

PART II: GENERATION 1 AND 2 OF THE EWQOS LOADING PROGRAM, ELPROG

19. The Environmental and Water Quality Operational Studies (EWQOS) Loading Program (ELPROG) was developed as an OE/RDBM system in SAS and consists of approximately 2500 lines of SAS procedure instructions and SAS programming statements. The impetus for the development of ELPROG rests in the biological, physicochemical, and research data management complexities involved with the pilot study phase of the long-term field studies. For instance, the four major areas of the project that deal with the various aspects of the ecology of the Mississippi River have a total of over 300 variables that potentially could have been measured. Therefore, because of the uncertainties involved in variable selection, sampling methodologies, analytical techniques, and the minimal effort that was allocated for data management, it was felt that no other approach seemed reasonable with as high a probability of success as the OE/RDBM approach.

20. The initial version of ELPROG was developed with data input being through a common Time-Sharing Option (TSO) file containing the card image data from all field measurements and laboratory analyses. ELPROG_1 inputs the TSO file supplying variable names and labels according to the type of data being processed (e.g., fish, benthos) and outputs these observations to a particular SAS data set, depending upon the type of data identified previously. This flow of data is shown in Figure 1. ELPROG_1 also made quality control decisions; when a variable for a given observation failed a quality control check, a flag variable associated with a particular variable was turned on for visual inspection of the printed SAS data set. The temporary, but saved SAS data set was edited using PROC EDITOR and/or UPDATE and usually, after several iterations, an error-free SAS data set was stored for processing by specific SAS analysis programs.

21. While the details of the loading program ELPROG_1 are beyond the scope of this report, several programming decisions were made prior to implementation of the system and are of interest. Briefly, ELPROG_1

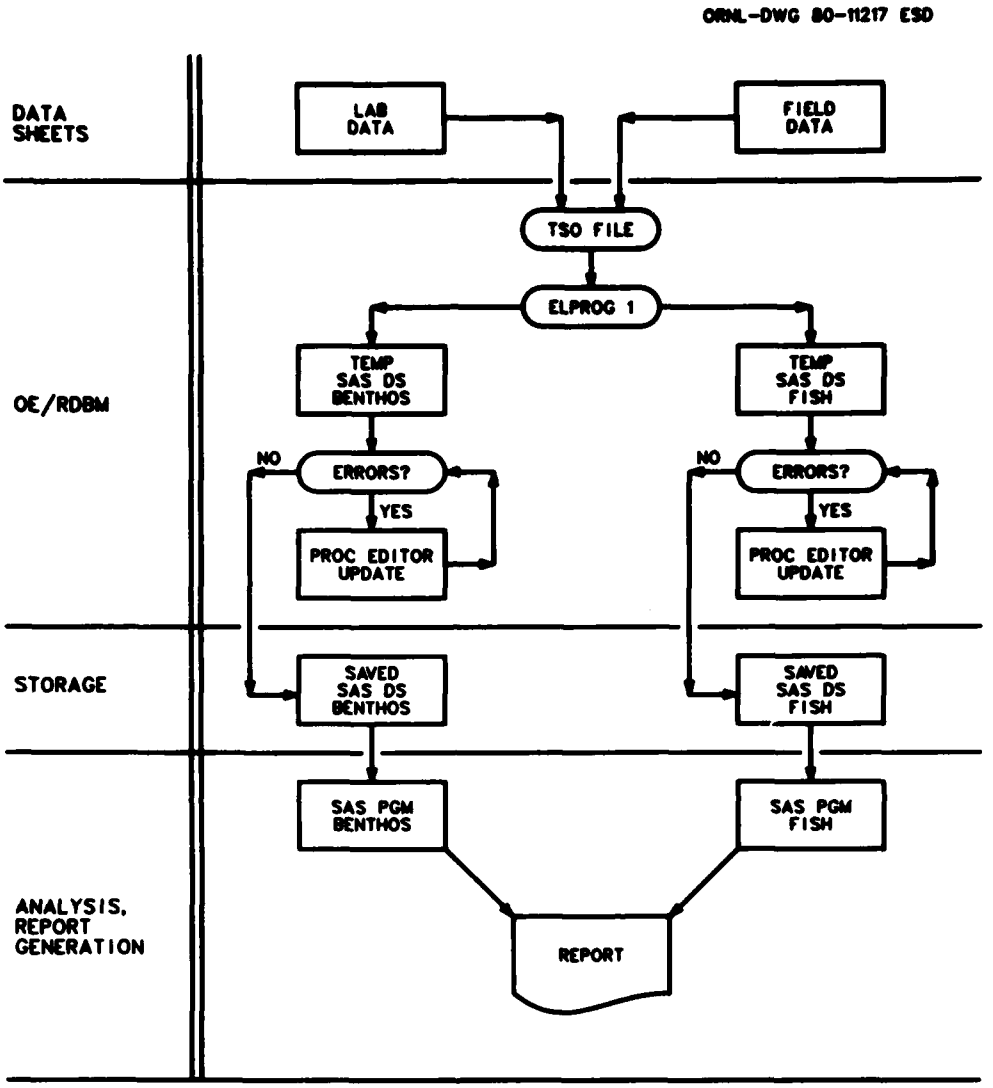


Figure 1. Data flow through ELPROG_1 for a particular period (Only two of the major work groups are shown (i.e., BENTHOS = benthic invertebrates and FISH = adult and larvae fish); DS = data set; PGM = program; OE/RDBM = open-ended research data base management.)

used a conditional input routine that keyed on a specific column for a unique character code to identify the input string and the appropriate data base for the observation. No alphanumeric codes were carried into the analysis programs. For example, all fish species codes were processed through a SAS MACRO that supplied the appropriate genus, species,

common name, etc. for the observation. Likewise, location codes, gear identification, time, dates, etc. were all translated into English equivalents.

22. During the first few months of the pilot study, variable lists changed on all input data sheets. As expected, however, the study followed a path that reflected the iterative nature of problem identification and solution. ELPROG_1 handled the many changes in the data base easily. Editing the temporary SAS data set for errors using PROC EDITOR and UPDATE was easily accomplished, but the amount of editing necessary to correct the data sets indicated that stricter controls were needed on data sheets prepared by the field crews.

23. Two major design problems were experienced during the longevity of ELPROG_1. The first problem was based on the assumption that the card image TSO file containing laboratory and field data would never have to be edited. In planning the OE/RDBM structure, it was originally thought that the saved SAS data set created after editing would be the only data set necessary for analysis and report generation. This assumption proved to be erroneous. The kind and complexity of logic errors experienced in the raw data file were not anticipated. The TSO file had to be edited or a SAS program had to be developed to handle all the encountered logic errors in the laboratory and field data sheets--a formidable task that would have conservatively taken 60 percent of the data management budget. With few other options available, editing under TSO/Multi Stage Variable (MVS) began. After the first major editing session, another problem emerged--the cost of editing under TSO. Complicating the TSO editing and increasing the cost of the editing was the fact that the new data files needing editing were random in relation to the position of an observation with erroneous information. Under these conditions, costs were astronomical (10-15 times greater) in terms of manpower and computer dollars and impossible to justify when data summaries and analyses might have to be reduced.

24. The second major design problem was the dropping of alphanumeric codes associated with many descriptive variables before the final SAS data set was saved. The overhead of carrying long,

alphanumeric variables along with each observation and the additional cost of sorting or merging on these expanded variables unnecessarily inflated data base management costs (see Part IV for discussion of coding).

25. It should be pointed out that the problems associated with ELPROG_1 were design problems and not an inability to meet new data base requirements by reprogramming the system. It is also felt that the OE/RDBM approach used in this project, with its limited planning activities, did not add to the probability of encountering the problems described above. On the contrary, the lack of any significant alteration of the data base structure and the ease with which the next version of ELPROG was implemented encourages the use of OE/RDBM using SAS.

26. The current version of the ELPROG system, ELPROG_2, has hopefully eliminated the problems experienced during the genesis of the initial system (Figure 2). One of the major changes incorporated into ELPROG_2 was the shift in the manner of handling errors in the data base. Major logic errors are now changed in the raw data file (TSO file) and a new SAS data set is compiled. For minor infractions, both the raw data file (TSO file) and the temporary SAS data set are modified, and the data flow continues to the storage of the final SAS data set. The other major change in ELPROG_2 is the substitution of expanded alphanumeric variables immediately before analysis or report generation routines. In this way, sorting and merging on numeric variables and/or short alphanumeric variables result in better timings, disk pack utilization, and cost. Storage requirements are also reduced because the overhead cost of carrying redundant descriptive variables with each observation has been eliminated.

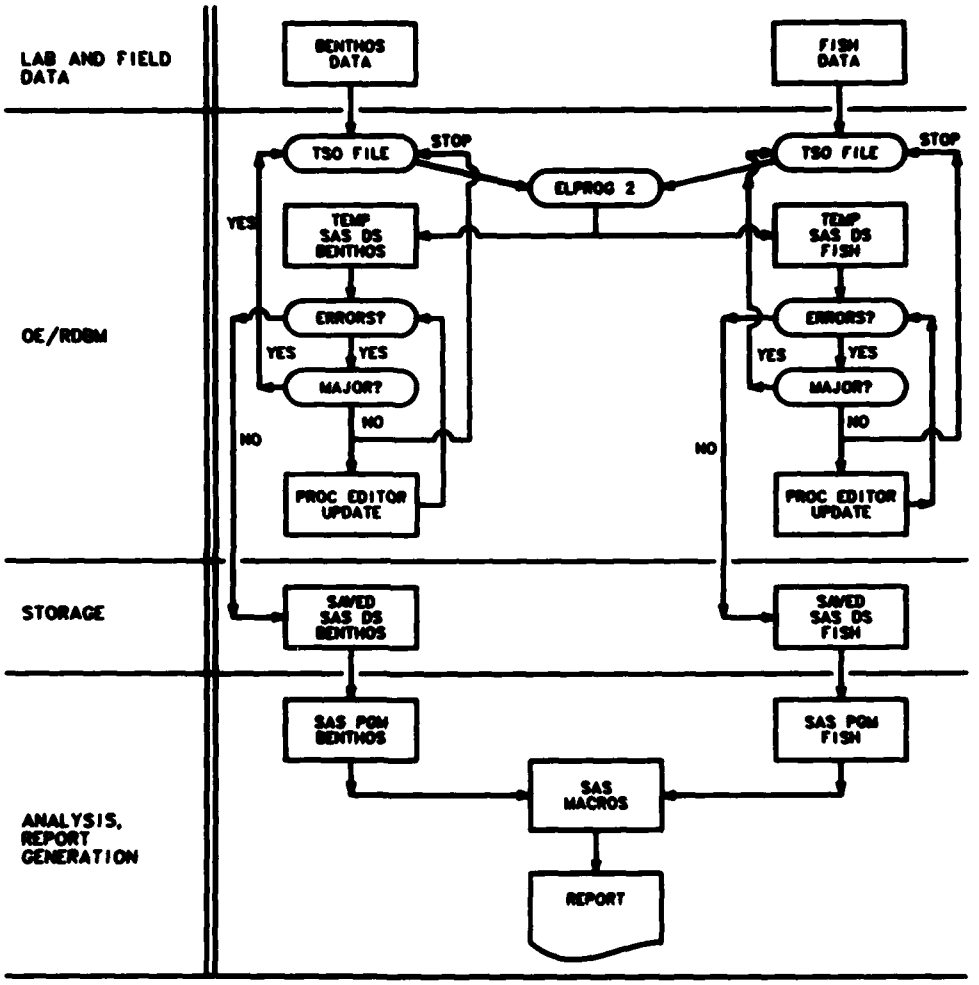


Figure 2. Data flow through ELPROG_2 for a particular sampling period (Only two work groups are shown (i.e., BENTHOS = benthic invertebrates and FISH = adult and larvae fish); DS = data set; PGM = program; MACROS = SAS program macro's; OE/RDBM = open-ended research data base management.)

PART III: GRAPHIC REPORT GENERATION

Alternatives for Graphic Report Generation

27. Three alternatives exist for copy-ready graphic report generation using SAS. The user had the following options: (a) develop and program his/her own specialized graphic routines for in-house use on locally supported hardware using the data management capabilities of SAS and incorporating the user's routines either as external programs to SAS or as SAS procedures; (b) use existing procedures supplied throughout the SAS Users Group, International (SUGI) SASware Index; or (c) purchase or lease graphic packages and interface with SAS (e.g., DISSPLA at Oak Ridge National Laboratories; Olson and Kumar 1980).

28. Depending upon need, frequency of use, the user programming ability, cost, and available monies, all three of these alternatives have met with varying degrees of success in meeting graphic requirements. However, each option has inherent problems in terms of applications to other projects and/or interest groups. In addition, cost/benefit considerations must be considered in most environments and especially when purchasing commercially developed graphic packages ($25,000-$50,000). Other areas of concern involve the programming skills of the developer; the efficiency of the routines not only in terms of CPU, core, disk pack utilization, etc., but also in the amount of personnel time needed for production; the lack of generalized routines that require little or no patches for each project application; and finally, the lack of hardware and software compatibility among the major graphic vendors with the subsequent need for redundant application programs being necessary for each type of plotting device.

29. Irrespective of SAS's current involvement in graphics, there may still be two main areas where graphic users experience difficulty. The first application would be the need for specialized graphical capablities of long- or short-term duration, but with limited funds and/or programming expertise available for development. The second area of difficulty involves those users with an established need for graphic

capabilities but with limited or no locally available hardware, or hardware systems that are either of low quality or are obsolete.

30. The impetus for the development of the first phase of the graphics system described in this paper reflects experience with most of the difficulties associated with copy-ready graphic requirements described previously. While the internal needs were quite specific and project-oriented, it was decided to develop a generalized graphic system compatible with SAS using a commercially available graphic package. Unlike the system developed at the Oak Ridge National Laboratories by Strand (1979), the graphic package VIVIDATA is a proprietary graphic display library of the McDonnell Douglas Automation Company (MCAUTO). Of paramount importance in the choice of VIVIDATA library was the package's ability to produce output files that may be postprocessed for use on a number of different hardware plotting devices. Hence, there is some degree of independence in terms of graphic hardware requirements.

The VIVIDATA System and Intermediate Plot Files

31. The VIVIDATA graphic display library is similar to CalComp and is designed as a FORTRAN-callable library. The library is modular in the sense that a particular task may be broken down into subtasks (e.g., drawing the axis and tic marks on an axis call) and is hierarchical in that, although lower level modules may be used directly, higher level requests may be made in which all or part of a task may be perfomed (e.g., a single call can produce a histogram). The package permits the user to specify the origin and units for any drawing area that can be related to a particular data value, or absolute in terms of the original page margins, or a combination of the two.

32. The VIVIDATA library exists in two interchangeable versions that form two different load modules for on-line and off-line displays. The on-line graphics terminal version supports displays on Computek and Tektronix storage tubes, while the off-line version, through the creation of an Intermediate Plot File (IPF) and postprocessing, supports displays on CalComp (748,663,960), Geber, Xynetics, and Houston pen and ink

plotters; Gould 4800 and Varian electrostatic plotters; and microfilm plotters such as the SC4460 or CalComp 835.

33. The basic differences between conventional and IPF plotting applications are shown in Figure 3. Instead of using a manufacturer's plotting software to develop a particular display program, the VIVIDATA graphics library is substituted for the usual call-type sequences in the program. The end-product of the compilation of the VIVIDATA instructions is a Binary Character Definition (BCD), 80-character logical record file containing vectors, characters, and control information. This file is the IPF which along with optional control parameters is input into another MCAUTO developed program, called a postprocessor, which decodes the records and produces a file to drive the desired plotter. The details of the VIVIDATA graphics library and the creation of the IPF can be found in two MCAUTO users manuals (McDonnell Douglas Automation Company 1979a).

34. Two points differentiate the VIVIDATA graphics system from other packages currently available. The IPF generation routines are plotter-independent and the need, therefore, for redundant application programs for different hardware devices is nonexistent. Secondly, the postprocessing programs are written in ANS level FORTRAN and have been developed for the IBM 360/370, IBM 1130, CDC 6000 series, Xerox Sigma, and Univac 1100 series. Hence, there is no computer dependence between the graphic system and the analysis and data management system.

PROC VIVIPLOT

35. As the first phase in development of a graphic system in SAS that will produce copy-ready figures, it was decided to program a SAS procedure in VIVIDATA very similar to the capabilities of PROC PLOT. Among other guidelines adopted in the development of the VIVIDATA-based system, the overriding consideration in all cases will be the use of a syntax that is as nearly identical to existing SAS procedures as possible. In the simplest case of the use of PROC VIVIPLOT, the user is only required to exchange procedure names, i.e., VIVIPLOT for PLOT, to produce

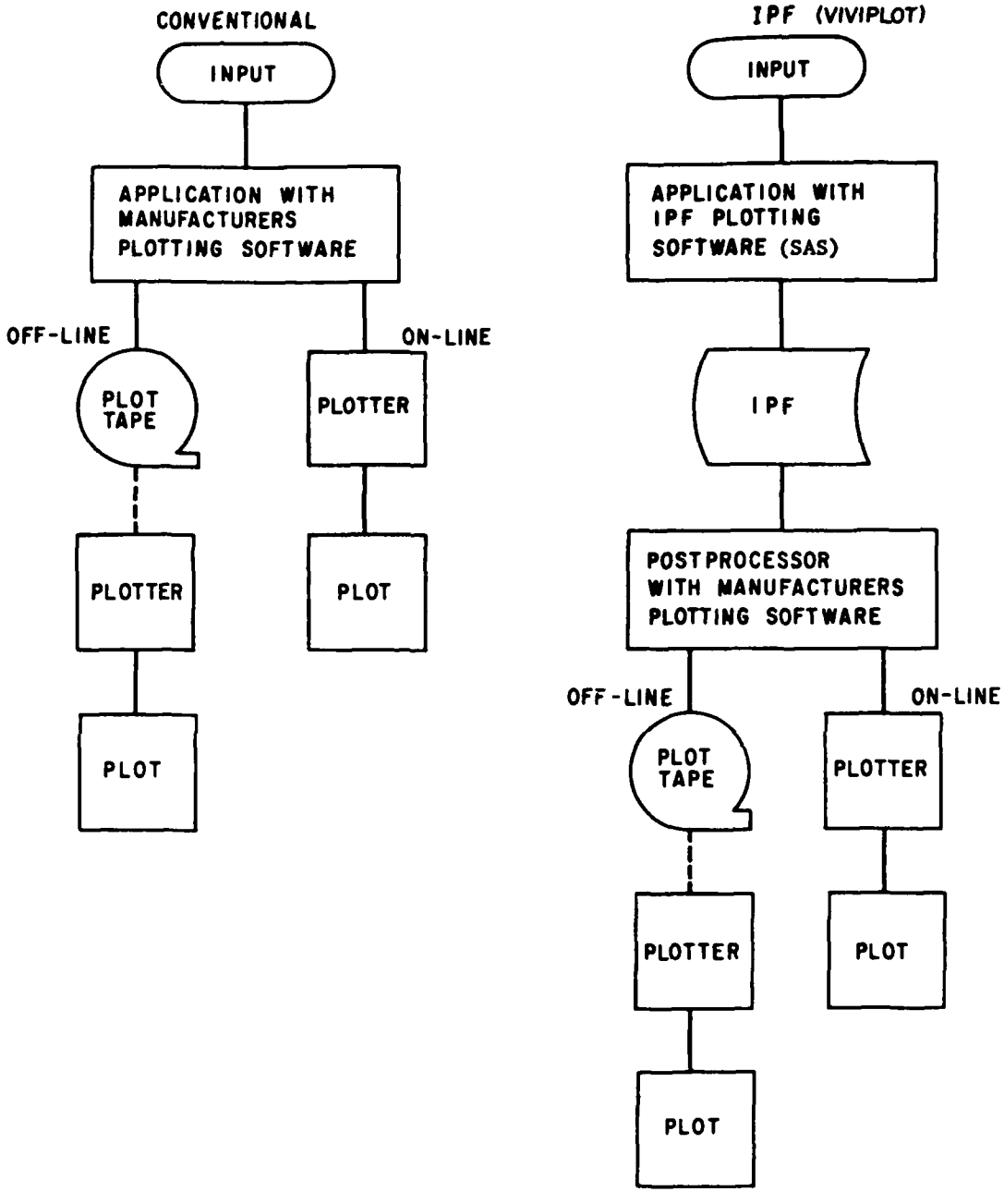


Figure 3. Plotting differences between a conventional system and a VIVIDATA-based system that produces an Intermediate Plot File (IPF) (The VIVIPLOT system is a SAS procedure written in VIVIDATA; figure adapted from Intermediate Plot File User Manual, McDonnell Douglas Automation Company 1979b.)

copy-ready two-variable plots. For example, if there is a SAS data set FISH with two variables--the number of larval fish per sample (FISH_NUM) and the velocity of the river at the sampling point (CURRENT)--a line printed plot is obtained by

```
PROC PLOT DATA=FISH;
PLOT FISH_NUM*CURRENT;
```

and a VIVIDATA drawn plot is obtained by

```
PROC VIVIPLOT DATA=FISH;
PLOT FISH_NUM*CURRENT.
```

36. The options and parameters that will be supported in VIVIPLOT's procedure and PLOT statement are shown in Table 2. The NOLEGEND option has been eliminated from the procedure statement since there will be only user-supplied headings and axis labeling (see VLABEL, HLABEL, and HEADING in Labeling Statements). The CONTOUR option will be replaced by another procedure devoted entirely to contouring. The parameters associated with the positioning and character labeling along the vertical and horizontal axes (i.e., VREF, HREF, VREFCHAR, and HREFCHAR) have also been eliminated and a grid system substituted. All other options and parameters of PROC PLOT including an identical syntax have been included in PROC VIVIPLOT.

37. The new option GRID in the PLOT statement provides the user with the capabilities of producing a grid system overlay on a plot along the marked coordinates of both axes. The symbols of the grid pattern are identical to those chosen for their respective axes.

38. The new parameter of the PLOT statement, CONNECT = variable $name_1$ (NOPRINT), variable $name_2$ (NOPRINT)...provides the user with the ability of producing plots with data values of selected variable names connected by a uniquely identifiable line (i.e., solid, dashed, dashed with triangles, etc.). The data values of a variable name may be printed as a user-supplied or program-supplied symbol or suppressed by the NOPRINT option immediately following the variable name. Up to 12 variable names may be listed as arguments of a CONNECT parameter of a plot request.

39. As an example of a VIVIPLOT plot with the data values connected and printed, a plot of the data values of FISH_NUM connected by a solid line and indicated by an asterisk would be produced by

```
PROC VIVIPLOT DATA=FISH;
PLOT FISH_NUM*CURRENT='*'/CONNECT=FISH_NUM;
```

If the data set FISH were to contain the predicted values (PRED) from GLM of the regression of FISH NUM on CURRENT, a plot of the data values of FISH_NUM as well as the predicted relationship indicated by PRED without the data values being printed would be produced by

```
PROC VIVIPLOT DATA=FISH;
PLOT FISH_NUM*CURRENT='+'
PRED*CURRENT/OVERLAY CONNECT=PRED NOPRINT;
```

A similar plot, but with the data values of FISH_NUM printed (+) and connected by a solid line, and the data values of PRED suppressed and conncected by a dashed line, would be produced by

```
PROC VIVIPLOT DATA=FISH;
PLOT FISH_NUM*CURRENT='+'
PRED*CURRENT/OVERLAY CONNECT=FISH_NUM PRED NOPRINT;
```

40. The capability of adequately labeling plots is mandatory in computer-based report generation. To this end, VIVIPLOT will provide the user with the ability to supply header or legend information, vertical and horizontal axis labeling, and reserved area identification and labeling on the plot itself.

41. On a VIVIPLOT two-variable plot, the user has the option of specifying up to 12 lines of alphanumeric legend or header information (LEGEND1, LEGEND2...LEGEND12) and/or labels associated with the vertical and horizontal area (VLABEL1, VLABEL2...VLABEL12 and HLABEL1, HLABEL2... HLABEL12). The syntax of the LEGEND, VLABEL, and HLABEL statements is identical to the TITLE statement, which is also supported in PROC VIVIPLOT. Like the TITLE statement, the alphanumeric labels of the VAXIS and HAXIS are centered along their respective axes, while LEGEND information is left-justified on the margin of the vertical axis. TITLE statement information precedes LEGEND information and is centered on the dimensions of the whole plot.

42. Taking advantage of VIVIDATA's concept of reserved areas within the axes of a plot, the users of PROC VIVIPLOT will have the ability of providing alphanumeric labeling within the coordinates of the axes. The position of the reserved box area without user-supplied instructions to the contrary will be determined by available space without hiding data values and scaling considerations. The user can, however, instruct VIVIPLOT to center the reserved box area in one of four possible positions, i.e., upper left quadrant (ULBOX), lower left quadrant (LLBOX), upper right quadrant (URBOX), and lower right quadrant (LRBOX). Reserved area box labeling syntax of up to 12 lines without choice of position is

```
BOX1 alphanumeric label;
BOX2 alphanumeric label; ...
```

To center the reserved box area in one of the four quadrants, the user substitutes ULBOX1, LLBOX1, URBOX1, and LRBOX1 and continues with the same syntax for other lines of label information (e.g., LLBOX1, LLBOX2, ...). The length of the reserved box area information cannot exceed 50 percent of the available positions along the vertical axis and is left-justified within the reserved area. Up to four reserved box areas may be requested for a single plot; however, in multiple reserved box area requests each box position has to be identified. Unlike the program-specified positioning of the reserved box area, user-specified quadrants may result in hidden data values. The reserved box area is outlined by the same symbols chosen for axes delineation.

43. As an example using the labeling statement and reserved box area capabilities of PROC VIVIPLOT, the previous example, using the data set FISH and the three variables FISH_NUM, CURRENT, and PRED, is used. To enhance the plot constructed previously, a reserved box area, legend labels, axes labels, and general title statements are added to the plot, i.e., the following instructions

```
PROC VIVIPLOT DATA=FISH;
PLOT FISH_NUM*CURRENT='+'
PRED*CURRENT/OVERLAY CONNECT=PRED NOPRINT;
LLBOX1 N=18;
LLBOX2 RSQUARE=0.96;
```

```
VLABEL1 LARVAE FISH DENSITIES;
VLABEL2 PLANKTON TOWS;
HLABEL1 LOCAL CURRENT;
HLABEL2 CM/SECOND;
TITLE1 ENVIRONMENTAL WATER QUALITY;
TITLE2 OPERATION STUDIES;
TITLE3;
LEGEND1 MISSISSIPPI RIVER;
LEGEND2;
LEGEND3 NEAR EUDORA, ARKANSAS;
```

would produce the plot shown in Figure 4.

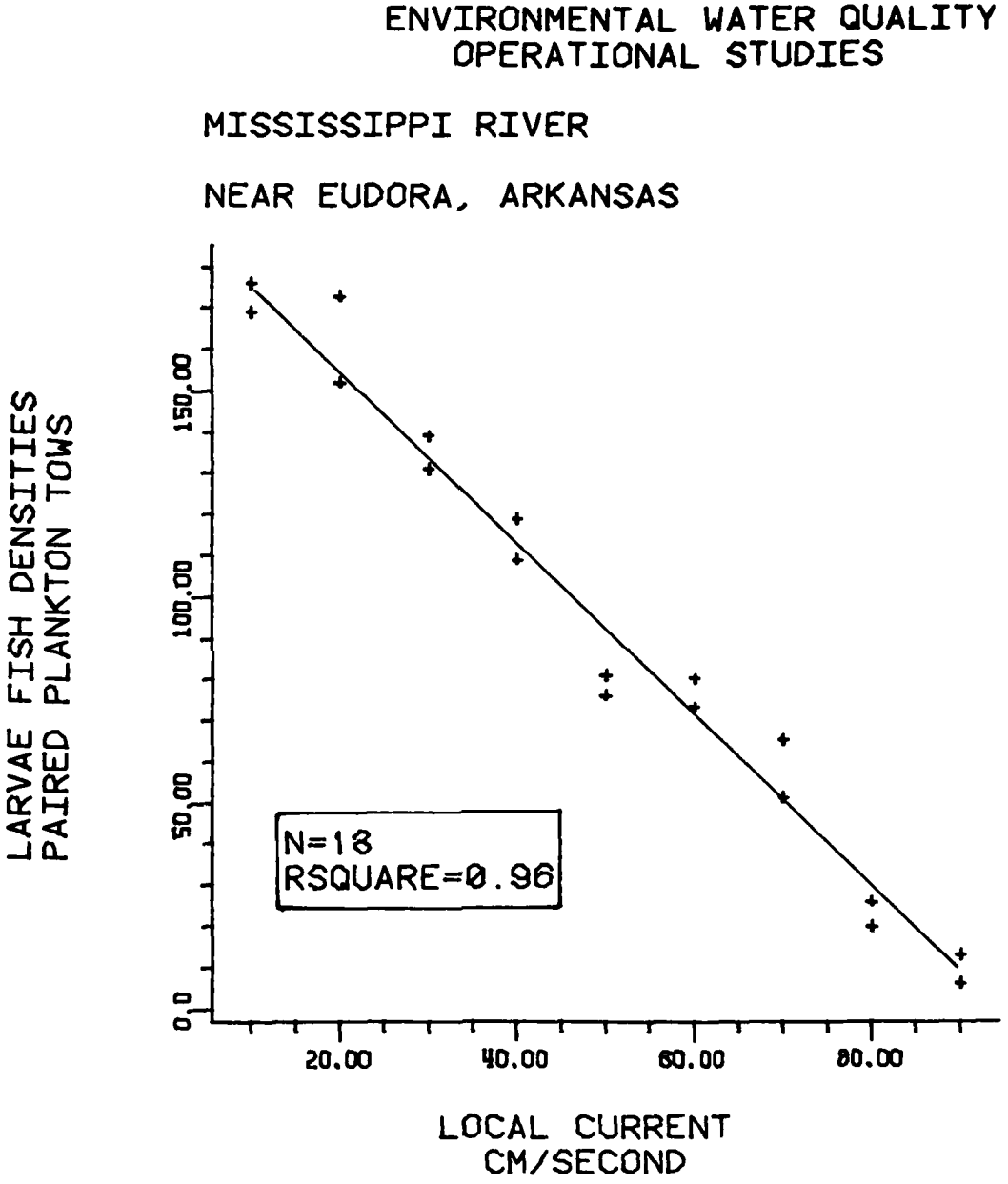


Figure 4. VIVIPLOT display of example

PART IV: NONSENSE CODES

Considerations in Coding

44. There is a great need to reduce the complexity of data encoding, increase storage efficiencies, and decrease error rates in complex studies using multiple data bases composed of hierarchical file structures. To meet these needs, RDMS have been developed with reduced error rates that emphasize the reduced complexity and the increased storage efficiencies needed in large, multidisciplinary studies.

45. Among the many problems found in integrating quality assurance controls into the RDMS are those associated with recording long alphanumeric sequences as abbreviated character strings or numeric codes. Numeric codes used to represent some precise terminology, such as chemical compounds, biological taxonomy, and geographic location, are quite common in RDMS (U. S. Army Corps of Engineers 1977).

46. Two specific problems have been observed when recording variable values. When alphanumeric sequences of even a few characters in length are recorded, it is virtually impossible to ensure proper and consistent coding of the value identifications. This is especially true when several data recording steps or recorders are involved in data processing where the amount of coding to be checked is excessive, and the codes, when they appear as abbreviations or in mnemonics, mean different values to various researchers. The second consideration is recording speed. When rapid sequences of coding of real-time events are required, a data recorder will not have sufficient time to manually record long alphanumeric event descriptors. The common solution to these problems has been, and probably will remain, the use of short codes to describe events, places, or things.

47. The purposes of this section are to point out potential difficulties associated with the use of codes with embedded information (i.e., "smart" codes) and to contrast storage efficiencies using "smart" codes with coding system in which no information is implied in the code (i.e., "nonsense" codes).

"Smart" Codes and "Nonsense" Codes

48. The use of "smart" codes which are either self-explanatory codes (e.g., M and F for male and female, respectively) or codes with embedded information contained in specific fields within the code itself is quite common in data sheet preparation. An example of an embedded smart code is the numeric string AcAcRu which has three two-digit fields each of which could represent biologically the family (Ac for Aceruceae), genus (Ac for *Acer*), and species (Ru for *rubrum*) codes of some taxon found in a sample.

49. In extant data bases, smart coding structures appear to be the most efficient means of sorting and subsetting data. However, in studies with developing data bases, there are at least two major considerations in determining whether or not to use a smart coding scheme. If the nomenclature system is simple (as in the male-female example above), the choice of short alphanumeric smart codes is obvious. However, as these codes expand to 10 or more characters, the probabilities of character transposition, misspellings, and actual storage costs become apparent. Reduced complexity in the data field is warranted. The second consideration when using smart codes is the knowledge of the universe of values of the variable being coded. If the universe is large but well-defined at the beginning of the program and the majority of the codes will probably be utilized, a smart coding system will function efficiently. For example, a large but well-defined universe of smart codes is used in the geoecology data base at Oak Ridge National Laboratory (Olson and Strand 1978; Olson, Klopatek, and Emerson 1979). The coded universe in this case consists of over 3000 county equivalents of the United States used to organize various environmental parameters. Each specific data base uses a five-digit numeric code to define what state (first two digits) and county/county equivalents (last three digits; U. S. Department of Commerce 1973) are being identified. This application of smart numeric codes to alphabetically sorted state and county equivalent values provides an efficient system for handling a large coded data base. The five-digit code is "smart" in that it has embedded information in the

code that was developed from a known universe of states and counties.

50. In situations such as those described above, the use of smart codes seems appropriate with rather straightforward error checking procedures. In situations like the male-female example, a simple conditional "IF-THEN-ELSE" programming statement can be used to check data entries. On the other hand, those applications exemplified by the geoecology data base can be quality assured by (a) conditional "IF-THEN-ELSE" programming checks, (b) evaluation of code frequencies within a given data set when the maximum allowable frequency of a coded value is known, or (c) merging of programs that list all codes not common between a reference data set containing proper codes and an experimental data set.

51. Smart codes, however, can be burdensome to the RDMS when either of the following conditions exists: (a) only a minor portion of the codes to be used is known (50 codes set up but only 5 used), and limited information is available as to the full classification scheme of specific values (5 codes set up but 50 needed), or (b) only a subset of the recorded codes will be utilized (50 codes set up, 50 assigned, but 5 actually used).

52. Under these conditions considerable time and money are spent developing a complex coding system that encompasses a large universe of entries and detailed classification schemes only to find little use of the majority of the codes established in the RDMS. In some instances, only 10 to 20 percent of the codes may be utilized. Nonsense codes, or codes without embedded information, efficiently circumvent the problems associated with smart codes. Using nonsense codes, alphanumeric variable values are assigned to a sequential numeric code as new values are encountered in the data base, irrespective of the position of the values in the classification scheme for that variable. The only knowledge required of the recorder is the last numeric code entered into the RDMS. When nonsense codes are used, the coding scheme is open-ended (Farrell, Magoun, and Daniels 1979, Strand 1979), and new variables and/or variable values are easily implemented.

Application of Computerized Nonsense Code

53. Nonsense codes were chosen for the EWQOS project, U. S. Army Corps of Engineers. The fisheries aspects of the waterway studies provide the background against which the decision was made to adopt the use of nonsense codes to describe the fish species encountered in any one of the three major waterways. Of the 2131 species of fish found in the U. S. (Bailey 1970), it was estimated that at a minimum the complete taxonomy for approximately 200 species would have to be coded to ensure that a high percentage of the species found in the river systems under study would be coded. Additionally, it was estimated that at least 10-15 ecological parameters (e.g., trophic level, commercial importance, growth stage, etc.) would have to be determined for each species entered into the RDMS in order to ensure appropriate categorizations for subsequent summary analyses.

54. In summary, there were two major difficulties in implementing a smart code approach for species definitions. Even if all 200 species were coded, there was still a high probability that several uncommon species would be encountered in the river ecosystems. Hence, the coding structure would require expansion to permit insertion of codes for these uncommon species. Leaving only the correct strategic gaps in the coding structure which permits the addition of codes is difficult at the beginning of a study. This approach seldom works in practice and usually increases (up to four characters in this case) the length of the code required to identify an entry. The second problem had to do with the difficulty of obtaining the proper values for each of the 10-15 ecological parameters associated with each fish species. A comprehensive literature search would have been necessary to ascertain some parameter values. Since it was already known that the majority of the species codes would not be used, the time and additional expenditures needed to determine these values could not be justified.

55. The solution to the fish species problem was to adopt a nonsense code approach in which a sequential numeric code beginning at 101 was given each species as it was discovered in any one of the rivers.

No adjustments were made in the coding scheme as to the taxonomic position of the species.

56. To date, 146 species have been collected and coded into the RDMS. Interestingly, at least 6 species were encountered that would have not been coded in the original 200 species. Inserting these species into a smart coding scheme might have been a minimal problem if the appropriate gaps were left in the code structure--an unlikely probability.

Error Reductions

57. While no statistics are available, the authors' experience suggests that studies employing nonsense codes have a reduced error frequency (10-30 percent reduction) over smart coding schemes. It is believed that one of the primary reasons for the reduction in errors associated with the use of nonsense codes might be the simplification of the recording process by removing correlative classification levels. Since no embedded information is contained in the nonsense code and classification levels are not correlated, there is no tendency for a code recorder to assume that two apparently associated values should code with approximately the same numeric code, when in fact the values are unrelated. Furthermore, no subject area skills are required of the recorder when using nonsense codes, since no hierarchical type of code identification is needed and only a simple accession number assignment is required.

Coded Output

58. Studies using nonsense codes or smart codes share two potential problems. From the end-user's viewpoint, if the computer output does not provide English equivalents for coded values, a separate code book must be kept to interpret the various computer codes. This approach is cumbersome and can introduce many subtle errors into the analysis and interpretation phase of a study. The second problem deals with those studies in which expanded values are generated for each value code

but are then needed to sort, merge, or subset the data base in order to produce the desired computer output. The cost involved with storing expanded variable values and processing these code labels with programs that sort, merge, etc. can be escalated using this approach and can seriously influence the capability of the RDMS to be cost-effective. With nonsense codes, the sorted order of the alphanumerics upon printing is the order of the nonsense code when sorted. This is one potential drawback of this code, but it can be alleviated through simple program statements in the SAS.

59. The FORMAT procedure in the SAS almost eliminates these problem areas and complements the nonsense approach using variable value labeling. Using PROC FORMAT to elaborate variable values provides a completely documented output from the RDMS and eliminates the need for a code book to interpret the computer output.

PART V: SUMMARY

60. The need for new methods in research data base management is acute. Historically, research data management has been based on planning and in-house development; however, this approach is becoming antiquated. New approaches in data management are needed to cure problems, but to be effective, these cures must be applied during the development of the data management system. The approach described in this report, open-ended research data base management, is intimately related to the capabilities of SAS. Under these conditions, OE/RDBM using SAS appears to be a practical and cost-effective alternative for managing large, complex research data bases. The practical experience gained from the implementation of ELPROG supports the usefulness of OE/RDBM and demonstrates the feasibility of the approach for research data management programs.

61. It is anticipated that users desiring both a flexible graphics system containing generalized SAS procedures and the ability to custom-tailor graphic needs to accomplish specific project goals will have potential use of the VIVIDATA-based system of which PROC VIVIPLOT is a phase one goal. With the current involvement of SAS in graphics, the VIVIDATA system is still extremely beneficial to those users identified in the introduction of this Part. Futhermore, the conceptual framework on which VIVIDATA is based (plotter and computer independence) makes this system a potentially significant contribution to the data base management. By eliminating the need for redundant application programs, a greater interchange of graphic routines would be possible among data managers. Furthermore, since existing CalComp-like graphic routines are easily converted to VIVIDATA, many device-dependent programs are now available for wider distribution.

62. With the development of the VIVIDATA-based system, the options for data managers with little or no graphic capabilities at their institutions are dramatically increased. These users can now transmit SAS data sets to MCAUTO computer centers and either produce copy-ready figures via PROC VIVIPLOT (presently estimated at $5-$12 for a plot of 50 observations with titles, legends, and axis labeling), or create an IPF for

off-line plotting on their in-house hardware.

63. Experiences in coding variables in a data management program have demonstrated increased efficiencies with nonsense codes versus smart codes with examples given that are being used to create output summaries for the Environmetrics program developed for EWQOS.

REFERENCES

Bailey, R. M. (ed.). 1970. "List of Common and Scientific Names of Fishes from the United States and Canada," Publication No. 6, American Fisheries Society, Washington, D. C., 150 pp.

Barr, A. J., Goodnight, J. H., and Saul, J. P. 1979. "SAS User's Guide," SAS Institute, Carrey, N. C., 494 pp.

Box, G. E. P., Hunter, S., and Hunter, W. 1979. "Statistics for Experimenters," John Wiley and Sons, N. Y., 653 pp.

Farrell, M. P., Magoun, A. D., and Daniels, K. 1979. "Management of Evolving Ecological Data Sets with SAS: An Open-Ended Management Approach," in: Proceedings, Fourth Annual SAS Users Group International Conference, SAS Institue, Carrey, N. C., 453 pp.

Martin, J. 1976. "Principles of Date-Base Management," Prentice-Hall, Englewood Cliffs, N. J., 352 pp.

McDonnell Douglas Automation Company. 1979a. "Intermediate Plot File User's Manual," M1956094, McDonnell Douglas Automation Company, St. Louis, Mo.

_____. 1979b. "VIVIDATA Reference Manual," M483087, McDonnell Douglas Automation Company, St. Louis, Mo.

Olson, R. J., Klopatek, J. M., and Emerson, C. J. 1979. "Regional Environmental Studies: Application of the Geoecology Base," in: Proceedings, Second Annual International User's Conference on Computer Mapping, Hardware, Software, and Data Bases, Cambridge, Mass., July 15-20, 1979.

Olson, R. J. and Kumar, K. D. 1980. "A SAS Graphics Procedure: DISPLA 1," Technical Manual 6993, Oak Ridge National Laboratory, Oak Ridge, Tenn., 39 pp.

Olson, R. J. and Strand, R. H. 1978. "Management of Diverse Environmental Data with SAS," in: Strand, R. H. and Farrell, M. P. (eds.), Proceedings, Third Annual Conference of the SAS User's Group International Conference, SAS Institute, Raleigh, N. C., 318 pp.

Strand, R. H. 1979. "Environmental Data: Management and Analysis Considerations," in: Proceedings, Fourth Annual SAS User's Group International Conference, SAS Institute, Carrey, N. C., 453 pp.

U. S. Army Corps of Engineers. 1977. "Dredged Materials Research Program," Technical Bulletin No. 1752, U. S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.

_____. 1978. "Information Exchange Bulletin, Environmental and Water Quality Operational Studies," Vol E-78-4, U. S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.

U. S. Department of Commerce. 1973. "Counties and County Equivalents of States in the United States," Federal Information Processing Standards Publication 6-2, U. S. Department of Commerce, Washington, D. C.

Table 1

Cost Analysis of an In-House Research Data Base Management System Compared to the Cost of an Open-Ended Research Data Base Management System (OE/RDBM)

Type of Work	Cost, Thousands of Dollars						
	Year					Totals	
	1	2	3	4	5	In-House	Commercial
In-House Development in FORTRAN							
Manpower*	55	55	55	55	55	275	275
Development†	15/30	15/30	15/30	15/30	10/20	70	150
Analysis	8/12	10/13	10/13	10/13	20/24	58	75
						403	500
OE/RDBM with SAS							
Manpower**	14	18	18	18	37	105	105
Development†	5/10	1/2	1/2	1/2	1/2	9	18
Analysis	8/15	10/15	10/15	10/15	20/30	58	90
						172	213

* Manpower cost based on a GS-12/1 skill level for a 12-month period at FY 80 costs.

** Manpower cost based on a GS-12/1 skill level for 3, 4, 4, 4, and 8 months respectively for years 1-5.

† Development cost for in-house computer/commercial computers.

Table 2

Relationship Between Existing Options and Parameters in PROC PLOT and the Procedure's PLOT Statement and the Anticipated Options and Parameters in PROC VIVIPLOT

Options and Parameters	VIVIPLOT Support
UNIFORM (By statement)	YES
NOLEGEND	NO
DATA=	YES
= variable_name	YES
CONTOUR	NO
OVERLAY	YES
VREVERSE	YES
	*CONNECT= **GRID
VAXIS=_____	YES
HAXIS=_____	YES
VPOS=_____	YES
HPOS=_____	YES
VSPACE=_____	YES
HSPACE=_____	YES
HREF=_____	NO
VREF=_____	NO
HREFCHAR=_____	NO
VREFCHAR=_____	NO

* = new parameter in PLOT statement of PROC VIVIPLOT.
** = new option in PLOT statement of PROC VIVIPLOT.

In accordance with letter from DAEN-RDC, DAEN-ASI dated 22 July 1977, Subject: Facsimile Catalog Cards for Laboratory Technical Publications, a facsimile catalog card in Library of Congress MARC format is reproduced below.

Aquatic habitat studies on the Lower Mississippi River, river mile 480 to 530 : Report 7 : Management of ecological data in large river ecosystems / by Michael P. Farrell ... [et al]. (Environmental Laboratory, U.S. Army Engineer Waterways Experiment Station). -- Vicksburg, Miss. : The Station ; Springfield, Va. : available from NTIS, 1982.
33, [2] p. ; ill. ; 27 cm. -- (Miscellaneous paper ; E-80-1, Report 7)
Cover title.
"February 1982."
"Prepared for Office, Chief of Engineers, U.S. Army under EWQOS Work Unit VIIB."
Bibliography: p. 33.

1. Aquatic ecology. 2. Computer program management. 3. Environmental engineering. 4. Mississippi River. 5. Water quality. I. Farrell, Michael P. II. United

Aquatic habitat studies on the Lower Mississippi : ... 1982.
(Copy 2)

States. Army. Corps of Engineers. Office of the Chief of Engineers. III. U.S. Army Engineer Waterways Experiment Station. Environmental Laboratory. IV. Series: Miscellaneous paper (U.S. Army Engineer Waterways Experiment Station) ; E-80-1, Report 7.
TA7.W34m no.E-80-1 Report 7