Naval Research Laboratory

Washington, DC 20375-5320

NRL/FR/5720--99-9912

Enhanced Emitter Identification Using Scaled Conventional Pulse and Intrapulse Parameters

SHAH-AN YANG
VIJAYANAND C. KOWTHA
GEOFFREY L. BARROWS
JOHN C. SCIORTINO, JR.

EW Support Measures Branch

DAVID A. STENGER

Center for Bio/Molecular Science & Engineering

September 8, 1999

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 8, 1999	Final

4. TITLE AND SUBTITLE

Enhanced Emitter Identification Using Scaled Conventional Pulse and Intrapulse Parameters

5. FUNDING NUMBERS

57-5320-A9

6. AUTHOR(S)

Shah-An Yang, Vijayanand C. Kowtha, Geoffrey L. Barrows, John C. Sciortino, and David A. Stenger

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/FR/5720--99-9912

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Security Agency
Ft. George G. Meade
MD 20755-6000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

Neural networks trained only on intrapulse (IP) parameters can yield a comparable level of emitter identification (EID) accuracy to what is currently achieved by experienced human analysts. We have extended this study using 314 collects from 42 emitters of the same model with the addition of three conventional parameters: pulse repetition interval, radio frequency, and pulse width. We examined the effects of pulse averaging using statistically derived distance measures. The two neural classifiers used outperformed the currently used matching algorithms. We find that using all three conventional parameters increases accuracy from 71.9% to 93.5%. Training using single-pulse representation was computationally expensive and revealed no difference in accuracy gained over pulse-averaged collects. Overall classification accuracy of greater than 95% is achieved using Mahalanobis distance measure. These results collectively show that EID can be significantly improved with the combined use of scaled unconventional and conventional parameters.

14. SUBJECT TERMS

Categorical knowledge
Emitter identification
Jack-knifing
Mahalanobis Distance
Neural networks
Pulse averaging
Pulse trains

15. NUMBER OF PAGES

42

16. PRICE CODE

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

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18
298-102

CONTENTS

EXECUTIVE SUMMARY

Previously,* we have shown that neural networks trained only on intrapulse parameters can yield a level of emitter identification (EID) accuracy comparable to what is currently achieved by experienced human analysts. In this report, we have extended that study by using three conventional pulse parameters (pulse repetition interval (PRI), radio frequency (RF), and pulse width (PW)) in addition to the intrapulse (IP) parameters. We performed this study using the same data set as that used in our previous work.* A test scenario was set up involving 42 different emitters, all of the same make and model, with multiple collects taken from each emitter under different circumstances. We first performed a statistical analysis to determine how reliable all these parameters were for EID. We then performed a variety of experiments to examine how much the accuracy of EID was improved by adding these conventional parameters. We explored both of these questions using the simple Euclidean distance metric as well as two other distance metrics: Euclidean distance with standardization, and Mahalanobis distance. We compared several different EID algorithms, including Nearest Neighbor and SuperPHC, the favored algorithms,* and two additional fielded algorithms. The experimental results suggested ways that better EID systems can be built, and are summarized as follows:

Both conventional and intrapulse parameters vary from collect to collect, but are reliable enough for EID. When the same emitter is collected under different circumstances, there is significant statistical variation in the collected pulses. However, the variation between collects is much larger between different emitters. Thus, the variation from collect to collect does not confuse EID algorithms. This observation is especially true when the parameter space is distorted in a statistically derived manner.

Using both conventional and intrapulse parameters improves EID accuracy. In all the experiments performed, using both the conventional and IP parameters resulted in greater EID accuracy than did using IP alone. The immediate implication is that baseline systems can have their automatic EID accuracy improved by incorporating conventional parameter information in their matching algorithms. This statement is especially true when the pulse data are standardized.

Averaging pulses over a collect before EID yields equivalent EID accuracy. Baseline systems work with pulse averages over entire collects rather than individual pulses. Our previous work* suggested the alternate approach for EID by first identifying every pulse in the collect, then taking a majority vote. Through experiments and statistical analysis, we demonstrate here that no significant advantage is obtained by using individual pulses for EID. Hence, the current baseline practice of working with pulse averages should be maintained.

Standardization of data is required for higher EID accuracy. When data are in raw form, some parameters have a larger variance than do other parameters. This is a result of the fundamentally different methods in which different parameters are obtained. For example, pulse width is measured by a clock signal while IP parameters are obtained from advanced signal processing. The magnitudes of the variances of each parameter are thus meaningless. However, when such raw data are scaled by using a Euclidean distance metric, those parameters with larger variance are given an unfair bias in the EID algorithm. To make the most of all the available information, the data need to be standardized before being given to an EID algorithm.

---

*G.L. Barrows, J.C. Sciortino, V.C. Kowtha, and D.A. Stenger, *Specific Emitter Identification Using Massively Parallel Implementations of Neural Networks*, Naval Research Laboratory Report NRL/FR/5720--96-9830, Sept. 1996.

Standardization with categorical knowledge results in extremely high EID accuracy. We observed EID accuracies in the 99% range when the data were scaled using categorical knowledge. This implies that in mission scenarios where the emitters have been previously seen, use of the techniques developed in this report will enable identification of these emitters with extremely high accuracy. The performance is not as high when there are many previously unseen emitters in a scenario. Nevertheless, this result implies that under certain types of mission scenarios, extremely accurate EID can be obtained.

Previous conclusions comparing SuperPHC and Nearest Neighbor are still valid when pulse averaging and standardization are used. In our previous work,* we concluded that Nearest Neighbor is slightly more accurate than SuperPHC, although SuperPHC is significantly faster in EID mode than Nearest Neighbor. We observed the same behavior in this study, whether standardization was used, whether pulse averaging over collects was performed prior to EID, or whether conventional parameters were used in addition to intrapulse parameters.

*G.L. Barrows, J.C. Sciortino, V.C. Kowtha, and D.A. Stenger, *Specific Emitter Identification Using Massively Parallel Implementations of Neural Networks*, Naval Research Laboratory Report NRL/FR/5720--96-9830, Sept. 1996.

ENHANCED EMITTER IDENTIFICATION USING SCALED CONVENTIONAL PULSE AND INTRAPULSE (IP) PARAMETERS

1. INTRODUCTION

Emitter identification (EID) is an important task in today's military environment. The goal of EID is to identify radio frequency (RF) energy sources by their pulse emissions. This is not a trivial task: identification must be made from libraries derived from pulse samples taken from thousands of different emitters. Most of these emitters are mass produced and differ only by serial number. Furthermore, the identification problem is made more difficult because the pulse samples that make up a library are taken under varying environmental conditions. Finally, mission demands require that EID be made within milliseconds, and be robust to noise and distortion [1]. Recent advances in receiver, signal processing, and pattern recognition technology have made significant progress in addressing these issues [2]. The different parts of an electronic support measures (ESM) system that perform EID have become sufficiently modularized to allow flexible development of specialized ESM systems.

As a result of ever more complex EW environments and reduced available human factors, there is a need to automate various parts of an ESM system. A modest goal is to automate the pattern EID tasks currently performed by an operator, thus freeing time to perform other tasks. A more aggressive goal is to automate the entire ESM system. Through the use of artificial intelligence techniques, it appears that such an autonomous real-time ESM tracking and analysis system is within the state of the art [2]. This requires the use of pattern recognition algorithms to deinterleave or separate pulse collects into individual emitter pulse trains. The deinterleaved pulse trains are then used to characterize the emitters for comparison with library data.

To illustrate how ESM systems can be automated, consider the nature of pulse data to be analyzed: the EID data consist of collections of pulses, each collection recorded by a single receiver over an interval of time. Ideally, the pulses collected would all be from a single emitter. Such an instance is called a "collect." However, depending on the exact filtering used and the current electromagnetic environment, a single collect can contain pulses from multiple emitters. Before advanced identification techniques can be performed, the incoming pulses need to be deinterleaved into separate sets representing different categories of emitters. This is a tedious task for human operators, and is thus a prime area for automation. Next a set of descriptors or parameters needs to be computed for each pulse category to allow comparison with previously collected or library data. Typically these parameters include the pulse repetition interval (PRI), which is obtained from the time of arrival (TOA), the carrier RF, the pulse width (PW) and, more recently, intrapulse parameters (IP). The problem of matching collected data with library entries is also tedious and susceptible to human error, thus is also a good area for automation. There has already been substantial work in using basic database matching techniques for identification. However, the application of more advanced pattern recognition and statistical techniques can significantly improve performance, as demonstrated in this report.

Previously, we have reported that by using only IP parameters, it is possible to automatically discriminate between emitters at an accuracy level comparable to skilled human analysts [3]. In that study, neural networks were given a training set consisting of 800 pulses from each emitter, then tested on "blind collects," which are defined as collects taken from previously seen emitters/modes, but not used in the training process. We performed identification on each individual pulse of a collect, and then used a majority-voting scheme to label the entire collect. This is different from the baseline approach of first averaging parameters over an entire collect and then identifying the "average pulse." By identifying every pulse in a blind collect,

Manuscript approved April 12, 1999.

and then taking a majority vote, the best algorithms were able to correctly identify the emitter source with greater than 80% accuracy. This study demonstrated the feasibility of automating the identification task. Recently, Sciortino et al. argued that using IP in conjunction with conventional parameters (RF, PW, and PRI) could increase EID accuracy into the 95% to 100% range [4-6].

In this report, we extend our previous study [3] by incorporating conventional parameters (PRI, RF, and PW), then challenging with data from the same field test. We also use two neural network classifiers (SuperPHC and Nearest Neighbor) that were found to be the most successful algorithms in the previous work. We again compare the performance of these two neural networks using the additional parameters. Finally, we compare the results of the study with other identification algorithms. Overall, this study provides a theoretical basis for analyzing and explaining the performance of both new and existing algorithms.

In baseline systems, the pulses are collected, processed, and then deinterleaved and identified as being from a single emitter. The pulse parameters from each single emitter are then averaged. The averaged features are entered into a library that forms a history of previously seen emitters. Effectively, the single "averaged" pulse parameter set from each collect is normally used for the identification task. In our previous work, we took the alternative path of using individual pulses over a collect instead of the pulse average [3]. In this report, we compare these two approaches (individual pulses and pulse-averaged collects).

The scaling of different parameters of data can also affect the EID accuracy [7]. If two parameters have the same amount of information, but one parameter has a greater variance over the other, then the importance of that parameter may be improperly exaggerated, resulting in sub-optimal EID. To address this problem, we standardized the data and used other scaling methods that standardize with respect to the variations within categories. We also explored the use of non-Euclidean distance measures to improve EID accuracy.

This report is organized as follows: Section 2 describes both the form of the pulse data used and the experiments performed. This includes methods of preparing raw pulse data for subsequent experiments. We then describe a number of EID accuracy experiments as well as a number of statistical tests performed on the data. In order to rationalize some of the experiments performed or steps taken, examples are taken from the pulse data. The overall purpose of this section is to acquaint the reader with the variety of methods of manipulating the data and performing experiments.

In Section 3, we explore the question of how reliable the pulse data are for EID. Several case examples are taken from the data sets. Then several statistical tests are applied to the data to assess their reliability.

Section 4 describes specific EID experiments and their results. First we describe the performance of EID schemes using both conventional and unconventional parameters. Then we explore how scaling and the use of Mahalanobis distance [8] can improve performance. A set of experiments is then performed to compare the performance of baseline systems with newer proposed approaches. This set of experiments validates both a) the pulse collection process and the subsequent EID as done currently in the baseline system, and b) the methods proposed in Refs. 3 and 4.

In Section 5, we discuss the experimental results and their significance. The implications for both future research and future systems are outlined. The two appendixes describe the raw pulse data in detail. This allows the reader to cross-reference data in described experiments with the raw pulse data. The date and time of each collect, location, a conventional parameter from ground truth, and labels used for category generation (Category and ID) are all presented for the two subsets (short- and long-range modes) used in this study.

2. BACKGROUND

This section describes both the form of the pulse data used and the experiments performed. First we describe pulse data in detail, including methods of preparing the data for experiments. We then describe a number of statistical tests used to analyze the data. Next we describe various distance measures that effectively distort the pulse parameter space according to pulse statistics. Finally we describe the EID algorithms used and the basic method for evaluating them. Examples taken from the pulse data are used to rationalize some of the experiments performed or steps taken. The overall purpose of this section is to acquaint the reader with the variety of methods of manipulating the data and performing experiments. We then provide a mathematical description of the ensuing experiments on scaling and categorical knowledge.

2.1 EID Algorithms

Nearest Neighbor and SuperPHC (Supervised Piriform Hierarchical Clusterer, Thuris Corporation, Irvine, California) are the primary EID algorithms used in this study. They have been described in detail earlier [3]. These algorithms are used in the majority of the experiments because they can be flexibly implemented to handle a variety of data representations. SuperPHC ran on a Sun Ultra1 Model 200 workstation. We implemented the Nearest Neighbor algorithm using Borland C compiler V5.0 (Borland International, Scotts Valley, California) on an IBM-compatible personal computer. The two other matching algorithms currently deployed are the Matching Algorithm [9] and the normalized Euclidean metric [10]. These methods are designed to operate on the intrapulse parameters. We did not attempt to fit the conventional parameters into the normalized Euclidean metric or the Matching Algorithm because those measures are not designed to be applied in that fashion. The Matching Algorithm and the normalized Euclidean metric generate match numbers, but we used them here only as a metric for a nearest neighbor network, which simplifies the algorithms. In this study, these algorithms are only used in the final comparison between baseline systems and the algorithms explored here (Nearest Neighbor and SuperPHC).

2.1.1 Nearest Neighbor

Previously, Nearest Neighbor was shown to be an effective algorithm for this EID task [3]. Nearest Neighbor is a simple, brute force, nonparametric method of pattern EID. In this algorithm, the training set vectors simply form a "prototype memory." Finding the closest vector in the prototype memory and assigning the unknown vector to the corresponding category identifies a test or unknown vector.

2.1.2 SuperPHC

SuperPHC is a biologically inspired neural network algorithm that is derived from simulations of the rat olfactory bulb and olfactory cortex [11,12]. It builds a tree to hierarchically partition the input vector space according to the statistical structure of the training set. If an individual category is located distant from other categories, a single partition may represent that category. If more categories are very close or overlapping, SuperPHC divides that area into finer and finer partitions until each category is separated. A new vector is identified by first locating it in one of the root-level partitions, then locating it into finer and finer partitions until the category of the vector is determined.

2.2 Data Description

The test data have been extensively analyzed [3, 13]. The objective of the test was to determine the effectiveness of current EID technology for identification and reidentification of small boat radar systems. A deliberate selection of many emitters having similar conventional parameters made it especially difficult to deinterleave and identify these emitters by conventional means. Most of the emitters used in the test were inexpensive, low-powered radars, 46 of which were identical models: the Raytheon R41X (Raytheon,

Framingham, Massachusetts). We used only the data collected from these systems for our tests. Thus the data used in this report simulates a difficult task very relevant to real-world environments. In fact, it can be said that the EID tasks simulated here are more difficult than what would be pursued in many mission scenarios.

Data sets consisted of two distinct parts, corresponding to different modes of operation for the emitters. Long pulses contain more energy and thus are used for long-range settings, while short pulses are used for local range settings. These two parts consisted of short- and long-range mode data. For the long-range mode data, there were 40 different emitters of the same make, differing only by serial number, comprising 152 collects of data taken during 12 days of the single emitter test. Note that the word "emitter" serves the same purpose as the word "category" in the context of this report. For the short-range mode data, there were 42 distinct emitters of the same make, comprising 163 collects of data. The pulses in the shorter-range setting tend to be shorter to allow more accurate determination of location. The pulses in the longer-range setting tend to be longer to allow more energy, hence range, per pulse. In these experiments, we treated the two subsets separately without loss of generality since data from the two subsets can be binned by size.

Each collect consisted of 896 pulses containing PW, TOA, RF, amplitude and intrapulse parameters. In this study, we used three conventional (PW, RF, and PRI) and 15 of the intrapulse parameters. Appendixes A and B list the data sets (42 and 40 categories respectively of short and long-range mode data) based on ground truth information [13]. In a prior study [3], emitters were found to be easier to recognize using long-range mode data than from the short-range mode data. This observation has also been made in another analysis of this data [13]. Rather than focus on just the long-range mode pulses, we chose again to use both sets because they contain pulses from most of the radars present in the test.

2.2.1 Removal of Outliers

The raw data sets contained stray pulses that needed to be filtered out before processing. In standard practice, one removes outliers by computing the mean of the set of data, then finding points farthest from this mean and removing them or substituting the outlying values with mean values or extreme values. It is appropriate to filter out these pulses before the ensuing experiments because these outlier pulses generally are instances in which severe multipath corrupts a single pulse or in which the radar antenna was not pointed directly at the receiver. Because the physical process that generates these outliers is significantly different from that which generates the statistical structure of the data cluster, and because these outliers are so easily removed, the analysis performed here is not compromised by removing these outliers.

To preserve the statistical information while removing outliers, we defined a neighborhood outlier removal function. As a distance measure, we used a Mahalanobis distance (described below) with a covariance matrix that is an average covariance matrix of all the collects. Within each collect (896 pulses), we found the Mahalanobis distance between each individual point and all other points in the collect. We found the N^{th} nearest neighbor to each point in the data, where N was the neighborhood size. The distance was noted for each point and then sorted, keeping track of each point and its associated distances. The points with the greatest distances associated with them were removed, leaving the desired fraction behind.

Experimentally, the removal of outliers had little effect on EID accuracy. It was only detectable in the pulse-by-pulse case, where it boosted the EID accuracy of individual pulses up to 10% but did not affect the EID accuracy after majority voting.

2.2.2 IP Parameters

Sixteen numbers represented the IP parameters. The first number is an offset term and not relevant to EID, so we ignored it. We used the remaining 15 IP parameters in our experiments.

2.2.2.1 Pulse Repetition Interval (PRI)

Before removing outliers from the data, we calculated the PRI. The process of extracting PRIs from the data begins by converting the time of arrival (TOA) data to delta time of arrival (DTOA) data. The TOA of a pulse is the time instant of the rising edge of the pulse. We discarded the first pulse, and for each following pulse computed the DTOA by calculating the difference between its time of arrival and the time of arrival of the previous pulse. Clusters were found in this distribution of DTOAs (when plotted on a histogram), and the locations of these clusters formed the PRIs. The clusters that were found in the distribution of DTOAs were either the PRIs or harmonics of the PRIs. Harmonics occurred when individual pulses were not intercepted and the resulting interval was the sum of two or more ideal DTOAs. The harmonics were then recognized by the fact that they are always more than twice the time duration of the shortest PRI. For simplicity, the harmonics were discarded. Figure 1 is an outline of the algorithm used to extract the PRIs.

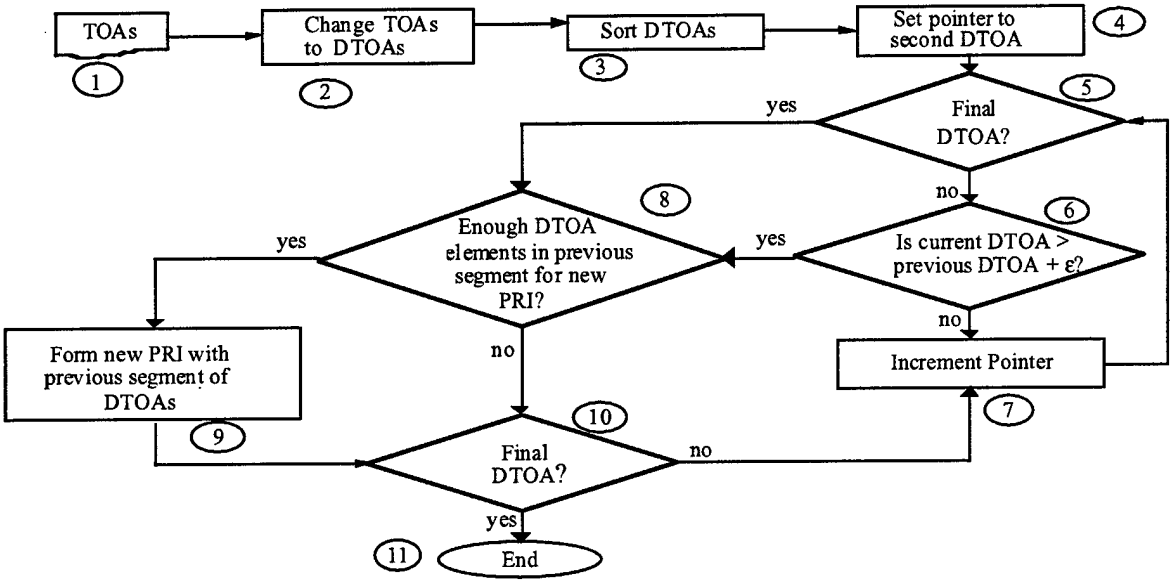

Fig. 1 — Flowchart for finding PRI from time of arrival (TOA) data

2.2.2.1.1 PRI Radius — The specific make of emitter used in this test had two PRIs for each mode of operation. For the short-range mode pulses, the two PRIs (legs 1 and 2) were approximately 340 μs and 520 μs. For the long-range mode pulses, the two PRIs were approximately 1100 μs and 1500 μs. The specific values of the PRIs varied slightly for each specific emitter. However, we observed that there was a linear relationship between these two PRIs (Fig. 2). In fact, the correlation between the PRIs exceeded 0.9999 and all PRIs were observed to follow this relationship:

$$PRI_2 = \frac{4}{3} PRI_1.$$

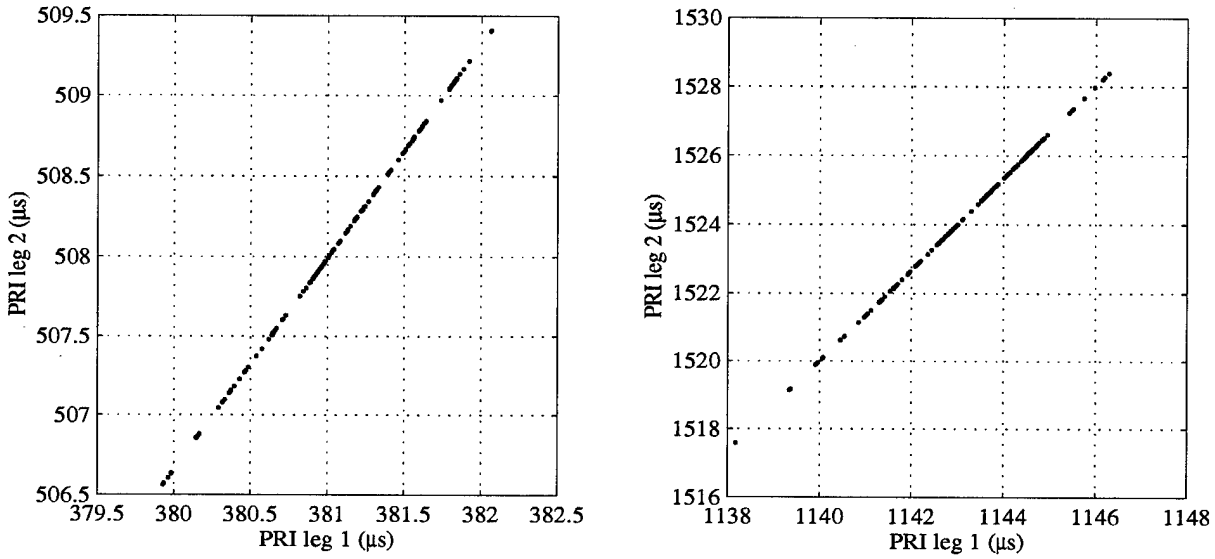

Fig. 2 — PRIs of the 42 emitters present in (a) the short-range mode (161 collect points overall) and 40 emitters present in (b) the long-range mode (153 collect points overall)

The close correlation between the PRIs would cause a covariance matrix that includes both PRI_1 and PRI_2 to be essentially singular. This prevents the use of both PRIs in a Mahalanobis distance metric. Therefore, we chose to represent PRI in the pulse-averaged experiments by a single parameter, the PRI radius:

$$PRI_{Radius} = \sqrt{PRI_1^2 + PRI_2^2}.$$

2.2.2.2 Radio Frequency

RF is a straightforward addition of the intermediate frequency (IF) term for each individual pulse and the tuned RF of the receiver. For more discussion on the computation of the IF term of each pulse, refer to Refs. 3 and 13. We defined the RF of each pulse as the sum of the receiver's frequency and the pulse's IF term. We used this reconstructed pulse RF as our data measurement for training and EID.

2.2.2.3 Pulse Width

PW, unlike the RF and the PRI, is a measurement that requires no special preprocessing. The PW is the time duration between the rising and falling edge (-3 dB points) of the pulse. We use PW directly in our EID tasks.

2.2.3 Statistical Analysis: H-test of Kruskal and Wallis

As described above, many collects from the same emitter can be obtained but with each collect taken under different environments. This is certainly the case with the MD-II data set (see Appendixes and Ref. 13). One important consideration is that of how much the statistics of each individual collect vary when taken under different circumstances and how this variation affects EID accuracy. To address this, we use the H-test of Kruskal and Wallis [14]. The H-test computes the likelihood that different clusters (here collects) of data were generated by the same statistical function [14-15], assuming this function is a joint Gaussian

distribution. In our case, we wish to examine the homogeneity of data taken on the same emitter, but in different collects. We hypothesized that different collects from the same emitter should statistically appear to be from a common population and therefore pass the H-test.

To perform the H-test, we first calculated the rank sums over the different samples by pooling the data from each sample into a common population. We then sorted the data and found the position or rank of each piece of data within that common population. Ties were resolved by averaging the ranks over which the ties occurred and assigning that average value to each element in the specific tie. We then separated the data into the different original samples and summed the ranks assigned to each data point over each sample, forming a rank sum for each of the original samples.

The variance of these rank sums formed a chi-square distribution with *N-1* degrees of freedom where *N* is the number of samples. In the H-test, we formed the following hypotheses:

H_0: Pulses from the same emitter from one collect to another come from the same population.
H_a: Pulses from the same emitter from one collect to another come from different populations.
H_0 and H_a are exhaustive.

Under the null hypothesis, H_0, the test statistic (shown below) has a χ^2distribution with *k-1* degrees of freedom. H_0 is rejected whenever $\hat{H} > \chi^2_{k-1;\alpha}$, where α is the significance level.

$$\hat{H} = \left[\frac{12}{n(n+1)}\right] \bullet \left[\sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3(n+1)$$

$$\hat{H}_{corr} = \frac{\hat{H}}{1 - \frac{\sum_{i=1}^{r}\left(t_i^3 - t_i\right)}{n^3 - n}}$$

where k is the number of samples;
R_i is the sum of the ranks of the i^{th} sample;
n_i is the number of data points in the i^{th} sample;
n is the total number of data points over all k samples;
t_i is the number of ties in the i^{th} tie; and
r is the total number of ties.

$\hat{H}_{corr}$ is a correction that should be used if ties occur more than 25% of the time, which will occur in our case because of the discrete nature of the data. In our analysis, we use one maximum significance level where the null hypothesis is not rejected.

2.3 Scaling and Mahalanobis Distance

This section deals with the different choices of scaling possible, and with using a distance measure other than Euclidean for the classifier. As a baseline, we use the simple Euclidean distance for EID. We then modify the distance measure by effectively distorting the pulse parameter space in a statistically derived manner that should improve EID. We discuss these methods in the following paragraphs. These methods include standardizing the data, using the Mahalanobis distance [12], using categorical knowledge, and finally variants that combine these methods. In all cases, the scaling factors and covariance matrixes are computed using the training data sets.

2.3.1 Standardization of Data

The process of standardization operates on individual parameters by rescaling each parameter so that the mean is zero and the standard deviation is unity. This removes the unfair bias in the distance measures given to parameters with a large variance. Each parameter is rescaled independently of the rest, then the simple Euclidean distance measure is used. Alternately, as described in the equations below, the distance measure itself is modified to generate the same result, with d_{ij}^2 the resulting distance measure:

$$\bar{\vec{x}} = \frac{\sum_{i=1}^{n} \vec{x}_i}{n}$$

$$\sigma_k^2 = \frac{\sum_{i=1}^{n} \left(x_{ik} - \bar{x}_k\right)^2}{n-1}$$

$$d_{ij}^2 = \sum_{k=1}^{M} \frac{\left(x_{ik} - x_{jk}\right)^2}{\sigma_k^2},$$

where n is the number of vectors;
M is the number of elements in each vector;
σ_{Li} is the i^{th} data vector;
σ_{Li} is the mean vector;
σ_{Li} is the k^{th} element of i^{th} vector;
σ_{Li} is the k^{th} element of mean vector;
σ_{Li} is the variance of k^{th} vector element;
σ_{Li} is the distance^2 between vectors i and j

2.3.2 Mahalanobis Distance

The Mahalanobis distance is a distance measure that stretches the parameter space according to the covariance matrix of the data set. Qualitatively, the Mahalanobis distance is similar to the standardized distance described above except that it is expanded to include second-order statistics between two parameters. In fact, if the parameters are independent according to second-order statistics, the covariance matrix is diagonal and the Mahalanobis distance gives the same result as the standardized distance above. Note that for the Mahalanobis distance to be used, the covariance matrix must be well behaved, i.e., invertible numerically:

$$\bar{\vec{x}} = \frac{\sum_{i=1}^{n} \vec{x}_i}{n}$$

$$C_x = \frac{\sum_{i=1}^{n} \left(\vec{x}_i - \bar{\vec{x}}\right) \times \left(\vec{x}_i - \bar{\vec{x}}\right)^T}{n-1}$$

$$d_{ij}^2 = \left(\vec{x}_i - \vec{x}_j\right)^T \times C_x^{-1} \times \left(\vec{x}_i - \vec{x}_j\right),$$

where σ_{Li} is the covariance matrix.

2.3.3 Use of Categorical Knowledge

We computed the individual scaling factors and the covariance matrixes over all pulses in all collects of the training sets. This method is not necessarily optimal because the locations of different pulse collects in parameter space dominate the statistics of each individual collect. Alternately, one can compute these statistics by including categorical knowledge. Qualitatively, this is performed as follows. For illustration, suppose our data set consists of two collects from the same emitter and we wish to compute the variance. Furthermore, suppose the pulses of each collect were generated using a Gaussian distribution. If we were to not use categorical knowledge, then we effectively merge these collects into one larger set, compute the mean of that larger set, then compute the variance from the distances between all the pulses and the resulting mean. The sum of distances squared, divided by the total number of pulses, would form the variance. If we were to use categorical knowledge to compute variance, then one mean for each of the collects would be first computed. The variance would then be computed by summing the squares of the distances between each pulse and the mean of the collect it came from, then dividing the sum by the total number of pulses in both collects. Thus, if the two collects were individually tight around their respective means, but the two means were separated, the computed variance using categorical knowledge would still be small. The computation of the covariance matrixes for the Mahalanobis distances is essentially the same.

2.3.4 Standardized with Categorical Knowledge

The first of the variants is called "standardized with categorical knowledge." In this scaling scheme, we took all the data sets and found all the emitters for which there were more than one collect of data taken. We then took these emitters and found the scaling factors of their parameters over their collects. Then we averaged the standard deviations over all emitters. Finally, we used these scaling factors as standardizing scale factors in our distance measure. Assume that we have a distribution of data with L categories. Each category will have a number of vectors, σ_{Li}. Let σ_{Li} denote the j^{th} vector of the i^{th} category. Also, let σ_{Li} denote the k^{th} element of the j^{th} vector of the i^{th} category. Let σ_{Li} denote the scaling factor for dimension k. Finally, let σ_{Li} denote the mean of the l^{th} category and σ_{Li} the k^{th} element of the l^{th} mean vector. Let σ_{Li} denote the distance between vectors i and j.

$$\vec{\bar{x}}_l = \frac{\sum_{i=1}^{n_l} \vec{x}_{li}}{n_l}$$

$$s_k^2 = \frac{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(x_{ijk} - \bar{x}_{lk}\right)^2}{\left(\sum_{l=1}^{L} n_1\right) - 1}$$

$$d_{ij}^2 = \sum_{k=1}^{M} \frac{\left(x_{ik} - x_{jk}\right)^2}{s_k^2}$$

2.3.5 Mahalanobis Distance with Categorical Knowledge

In our second variant, Mahalanobis distance with categorical knowledge, we first found the scaling covariance matrix for each category of data. Next, we calculated an average of this covariance matrix and

used it as a general covariance matrix for describing all of the collects. This covariance matrix was used in Mahalanobis distance computations. The notation is the same as above, with σ_{Li} being the scaling matrix.

$$\vec{m}_l = \frac{\sum_{i=1}^{n_l} \vec{x}_{li}}{n_l}$$

$$SC_x = \frac{\sum_{l=1}^{L}\sum_{i=1}^{n_l}\left(\vec{x}_{li} - \vec{\bar{x}}_l\right)\times\left(\vec{x}_{li} - \vec{\bar{x}}_l\right)^T}{\left(\sum_{l=1}^{L} n_l\right) - 1}$$

$$d_{ij}^2 = \left(\vec{x}_i - \vec{x}_j\right)^T \times SC_x^{-1} \times \left(\vec{x}_i - \vec{x}_j\right)$$

2.3.6 Generalization of Scale Factors

Our practice here of using all the data to generate the covariance matrix for our classifier is unrealizable in practice because we will not know categories before we identify the data, although it was necessary in our previous experiment. In reality, only a fraction of all possible emitters can be used to form such covariance matrixes or scaling factors because new emitters will appear and their statistics cannot be known a priori. Therefore it is necessary to study the effects of using such categorical knowledge when it is derived from a subset of all the collects. This is studied in detail in Section 4.

2.4 Matching Algorithm

Since we were only using a nearest neighbor network instead of generating a match number, the matching algorithm simplifies to

$$d^2 = \sum_{i=1}^{15}\left(\frac{v_i - v_{Li}}{\sigma_{Li}}\right)^2 \times \sum_{1}^{15}\sigma_{Li}^2,$$

where σ_{Li} is the standard deviation of the library coefficient, σ_{Li} is the mean of the i^{th} library coefficient across all signals used to make the entry, and σ_{li} is the mean of the i^{th} coefficient across all signals in the collect [10]. We simply used d as our distance measure in our nearest neighbor network.

2.5 Normalized Euclidean Metric

Again, we only used this metric as a distance measure for our nearest neighbor network. The notation is the same as above in Ref. 6.

$$d = \frac{\sqrt{\sum_{i=1}^{15}\left(v_i - v_{Li}\right)^2}}{\sqrt{\sum_{i=1}^{15} v_i^2} + \sqrt{\sum_{i=1}^{15} v_{Li}^2}}$$

2.6 Jackknife Evaluation

The EID accuracy experiments were conducted using the "jackknife" network strategy. Jackknifing is useful for evaluating an EID scheme. In jackknifing, we remove a single data point from the data set and build a classifier using all the remaining data. We then use the resulting classifier to identify the original data point that we removed. The experiment is repeated for all points in the data set. The main advantage of jackknifing is that the EID accuracy obtained is less dependent on the specific collects used to form the training and testing sets in individual trials.

3. STATISTICAL RELIABILITY OF CONVENTIONAL AND INTRAPULSE PARAMETERS

One question that needs to be addressed is the statistical reliability of the parameters used for EID. Clearly within each collect there is statistical variation of the pulses due to the random noise that corrupts all signals. However, statistical variation between different collects of the same emitter has been observed. These variations occur because such collects were typically collected under different circumstances and environmental conditions. The important question is whether this variation is significant enough to prevent accurate EID. This section addresses this question by examining the relevant statistics. First we look at several cases of variation between different collects of the same emitter. Then we perform the H-test on the grabbed data to determine if it is statistically likely that such different collects came from the same emitter. Finally, we compare the variation between collects of the same emitter with variation between collects of different emitters.

3.1 Case Examples

First we consider examples of how the statistics vary among different collects of the same emitter. Figure 3 shows the relative distribution of two intrapulse parameters (A1 and B1) of 43 collects from two different emitters (categories 9 and 33 from short-range mode collects; categories 1 and 8 from long-range mode collects). Individual collects form a category-dependent cluster in the A1-B1 space. However, at least one category 8 collect overlaps with 18 category 1 collects (Fig. 3, left panel). This overlap would cause

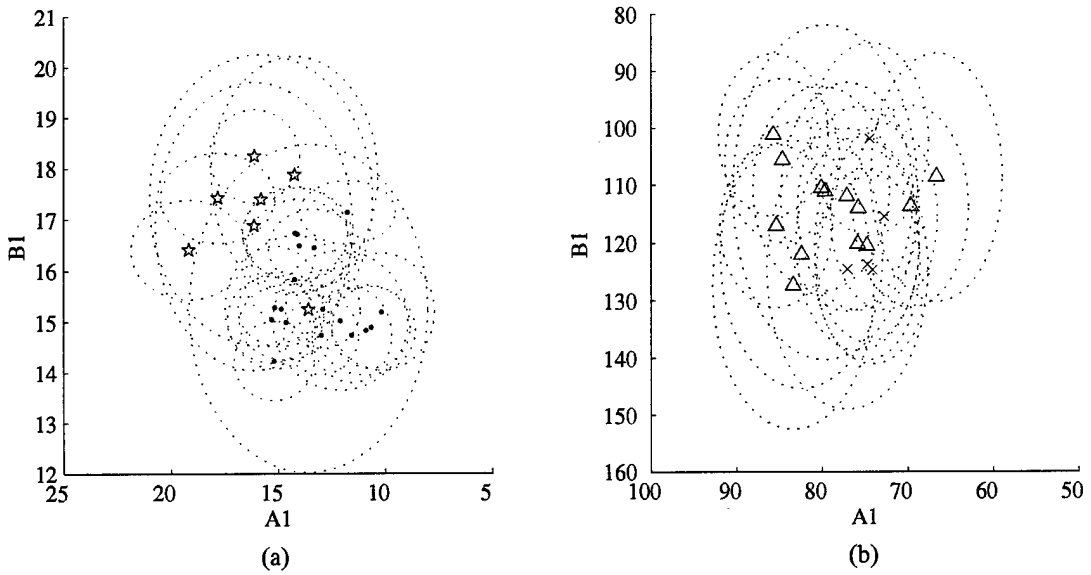

Fig. 3 — Distribution of two intrapulse parameters (A1 and B1, mean and one standard deviation represented by four symbols and dotted lines, respectively) of 43 different collects for (a) long-range mode data (18 collects from category 1 and 7 collects from category 8), and for (b) short-range mode data (13 collects from category 9 and 5 collects from category 33)

errors in an EID algorithm. The overlap is severe in the short-range mode data (Fig. 3, right panel): note that all category 9 collects fall within category 33's cluster space. An emitter identification scheme acting on this data would not be able to discriminate these two emitters using these parameters. Note that parameters A1 and B1 shown in the figure are generally regarded as the most reliable parameters for emitter discrimination [4]. This observation is supported by an earlier study, in which we showed with information theory that these parameters are among the best of the intrapulse parameters for EID [3].

Figure 4 presents an example of how there can be variations between different collects of the same emitter. This IP variation is also evident from two collects of the same emitter taken minutes apart. Figure 4 shows an example of A1 parameter variation from one of the above emitters (category 1, long-range mode data, Appendix B) from two different collects that were collected 94 min. apart from the same site. The first collect has a mean of –15.1 for A1, but the second collect has a mean of –12.0 for A1. The means are actually more than one standard deviation apart from each other using the standard deviation from either collect. These means are statistically different even at the 0.001 significance level according to the T-test. Such variation is not limited to IP coefficients. Figure 5 shows changes over the 12-day MD-II test period in five intrapulse parameters and three conventional parameters. These plots were generated using the same category 1 data in the long-range mode data shown above in Figs. 3 and 4. These plots show that the distribution of both sets of parameters varied during the 12 days of the MD-II test period. The variations are prominent in the first two days of collects and stabilized with time.

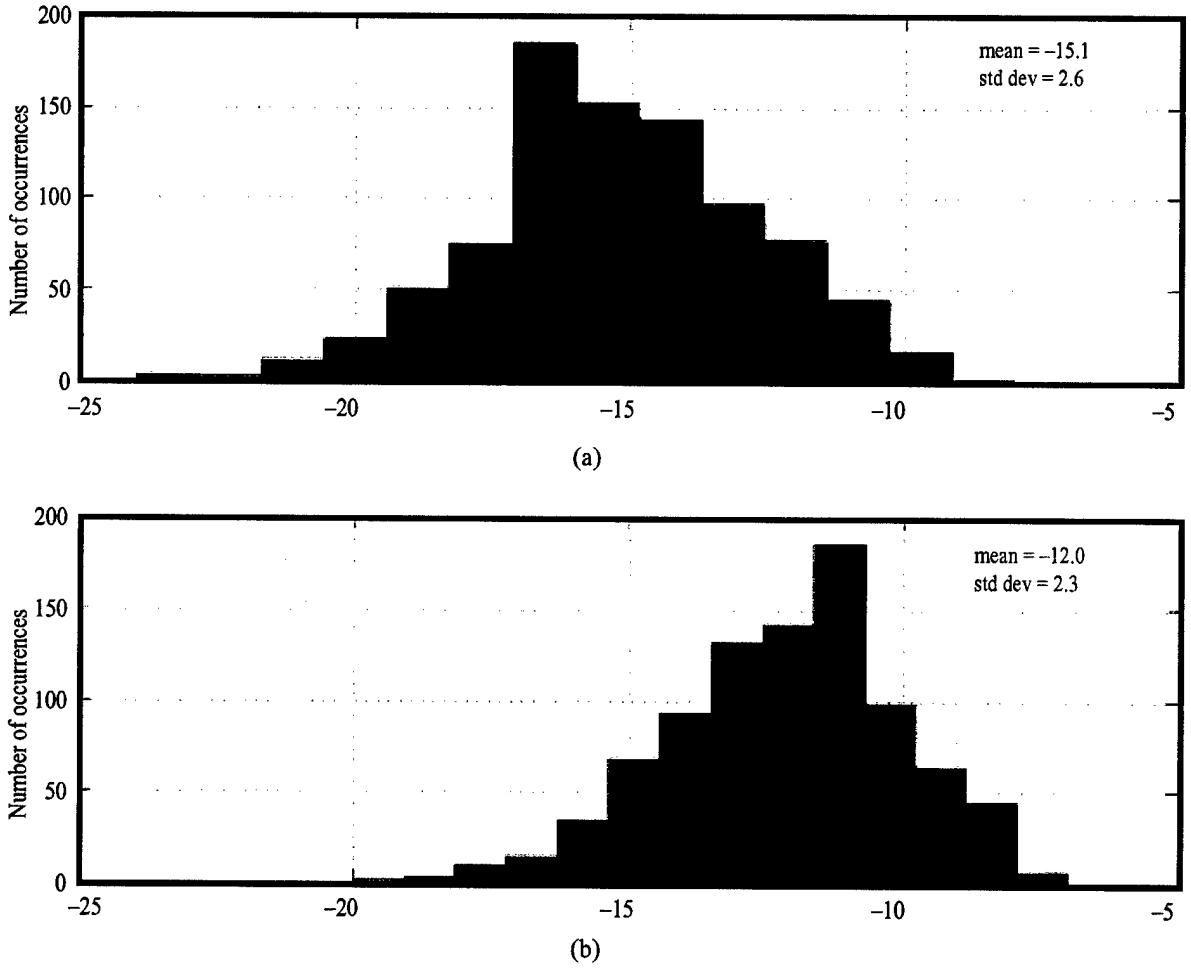

Fig. 4 — Two histograms of parameter A1 of an emitter (>800 pulses each) from long-range mode for (a) category 1, collect 2, and (b) category 1, collect 16, from the same site recorded 94 min apart on day one of the test

3.2 H-test of Kruskal and Wallis

Figure 5 shows examples of how the conventional and intrapulse parameters might vary between emitters. Here we perform a more quantitative analysis using the H-test, which is described in Section 2.2.3. The H-test looks at multiple distributions of data points and computes the likelihood that these distributions come from the same "population," or from the same emitter in the context of this study. The H-test works by ranking the data from the different samples and comparing the average ranks for each sample. We expect that if the collects were from the same population that the ranks would be fairly evenly distributed over the different collects.

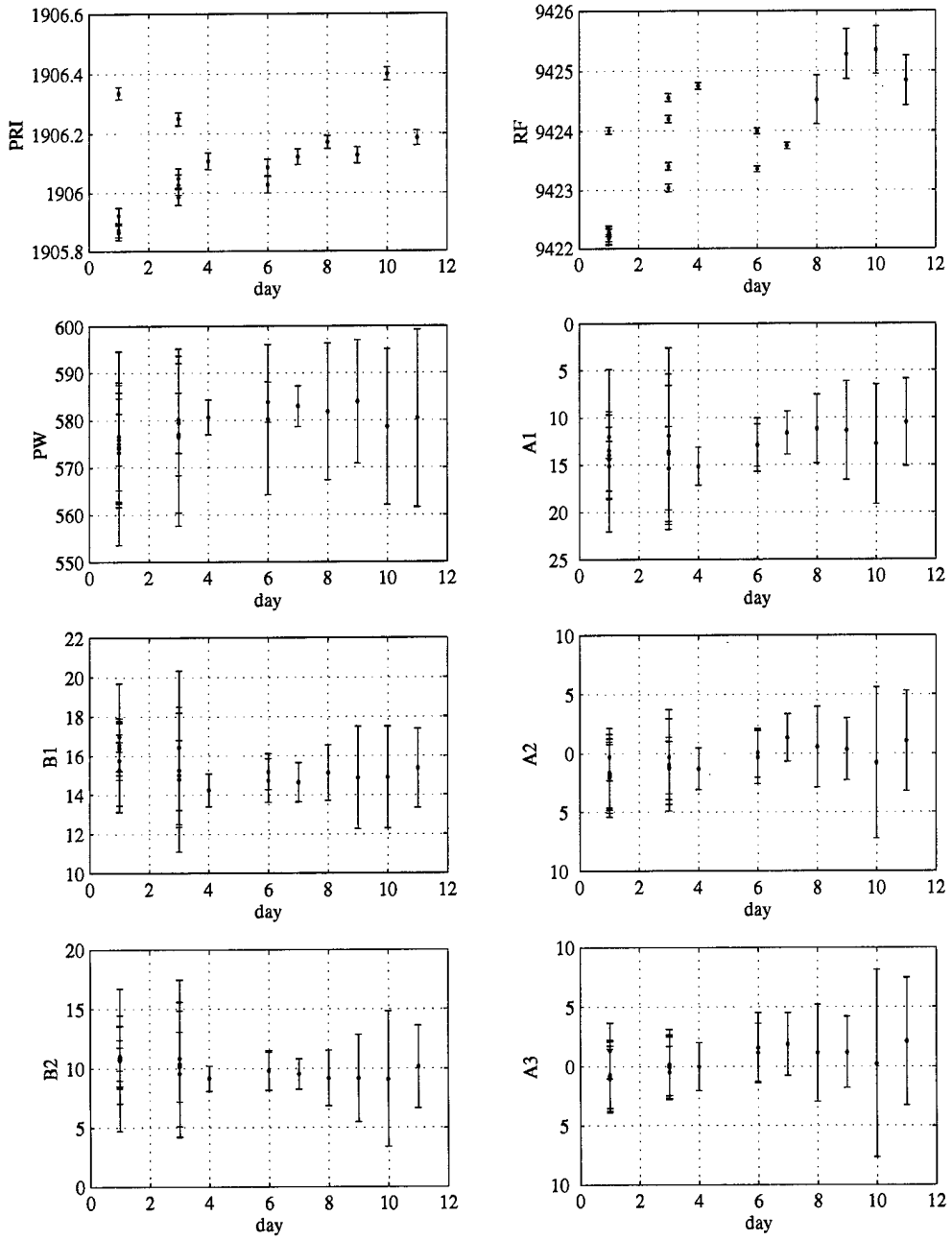

Fig. 5 — Variation of PRI, RF, PW, and five intrapulse parameters (A1-A3, B1-B2) with time in the 18 collects made (1-12 days of category 1, long-range mode, Appendix B)

Table 1 shows the probability distribution using H-test for the 17 parameters (RF, PW, and 15 IP parameters for both short-range and long-range data) studied. The probability value of ~1.0 indicates that the alternate hypothesis is correct (i.e., pulses from the same emitter from one collect to other come from different populations). With an overall average of 96.9% certainty, the distributions formed by the pulses of different collects of the same emitter do not appear to be samples from the same emitter. These results imply that pulses from different collects of the same emitter do not form a single homogeneous population in a statistical sense. It should be noted, however, that further analysis shows that this intraemitter variation (examined below) is significantly less than the variation between emitters. A secondary implication is that no significant advantage is obtained by using single-pulse data. Thus, it is just as effective to use pulse averages over collects of the conventional and intrapulse parameters as it is to retain the original pulses. Therefore, a single data point can be used to represent an entire collect without sacrificing EID accuracy.

Table 1 — Probability Distribution Using H-test

	Long-Range	Short-Range
RF	>0.999	0.9965
PW	0.9635	0.9966
A1	>0.999	0.9956
A2	0.9790	0.9823
A3	0.9659	0.9894
A4	0.9548	0.9817
A5	0.9898	0.9801
A6	0.9415	0.9735
A7	0.9400	0.9924
A8	0.9581	0.9836
B1	0.9985	0.9940
B2	0.9845	0.9980
B3	0.9439	0.9907
B4	0.9409	0.9934
B5	0.8686	0.9636
B6	0.9058	0.9657
B7	0.8958	0.9918

3.3 IP Parameters Distance Variation Between Different Emitters

In this section, we compare the intercollect variation within emitters with the variation between emitters. Qualitatively, if the variation of collects between the same emitter is less than that between different emitters, then the pulse data contain enough information for identification. This study is performed as follows: first we compute all of the distances between the different collects. Then the distances resulting from different collects of the same emitter are placed in one set, while the distances resulting from different collects of different emitters are placed in a second set. Overall, the 314 collects from short- and long-range mode data yielded 423 and 444 same-category distances and 12,618 and 11,032 different-category distances, respectively. These two sets of distances are then each plotted in a histogram. All three distance measures used throughout the report (standardized scaling, standardized scaling with categorical knowledge, and Mahalanobis distance with categorical knowledge) are evaluated using the following formula:

$$d^2 = \sum_{i=1}^{m} \left(\frac{\mu_{1i} - \mu_{2i}}{\sigma i} \right)^2 .$$

For standardized scaling, μ_i and σ_i are averages and standard deviations, respectively. For standardized scaling with categorical knowledge, the σ_i is the diagonal of the covariance matrix. Finally, for Mahalanobis distance with categorical knowledge, the σ_i is formed from the inverse covariance matrix. The percent overlapping distance between the same and different collects is then computed. For clarity, only the number of instances of different emitters overlapping with same emitter collects is considered here. Next, a percentage of overlap (number overlapped as a function of more than 10,000 different category instances) is computed for all the metrics used for the two subsets of data.

Figure 6 plots the variations in IP distance between (a) the same and (b) different category collects for long-range mode data. The computed distance was greater than 1.8 in 36 instances of same category collects and was less than 1.8 in 970 instances for different category collects (less than 9% overlap). Thus, the computed IP distances between collects of the same category and collects between different categories form two different distributions with a small overlap. This formation of two different distributions for the same and different category collects suggests that IP parameters can be used to effectively discriminate between emitters. The overlap for the short-range mode data was twice as high (more than 20%) for different category collects (figure not shown). For short-range mode data, EID using 15 IP parameters was less than 60% (see Section 4). This is due to a larger overlap (greater than 20%) between the same and different category collects. Note that the overlap is reduced with scaling and is discussed elsewhere (see Section 4).

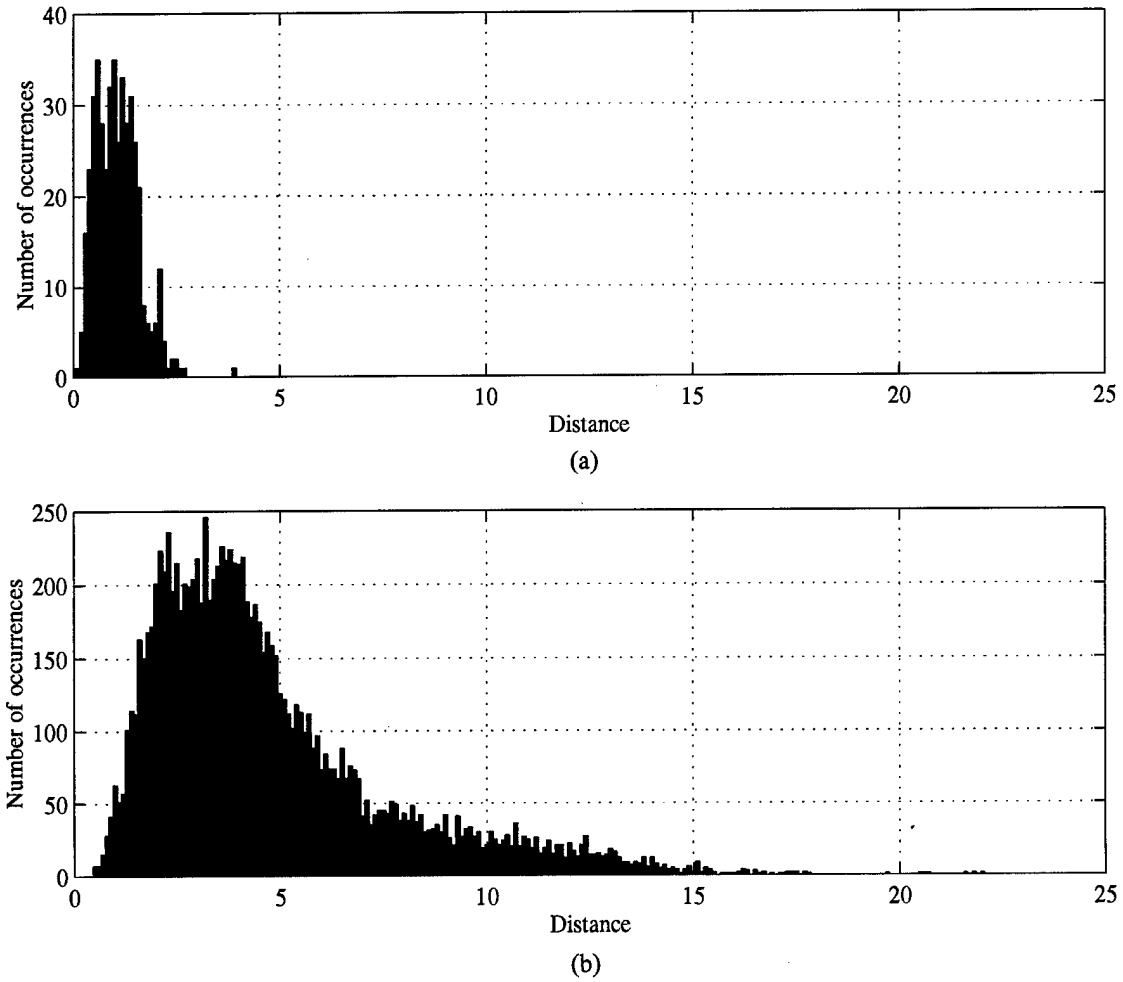

Fig. 6 — Distribution of IP distance between same category collects (a) and different category collects (b) for standardized long-range mode data (152 collects, 444 same category variants, and 11,032 different category variants)

Figure 7 illustrates the overlap between intraemitter and interemitter distances. The overlap is higher in the short-range mode data for all three types of scaling used in the present study. However, the overlap drops off significantly with scaling. Note that when using standardization with categorical knowledge, the overlap falls below 1% but is never 0%. The overlap drops further when the Mahalanobis distance with categorical knowledge is used. Thus, the collects vary in intrapulse and conventional parameter space but form distinct distributions with minimal overlap. These results are consistent with the EID accuracy results obtained above.

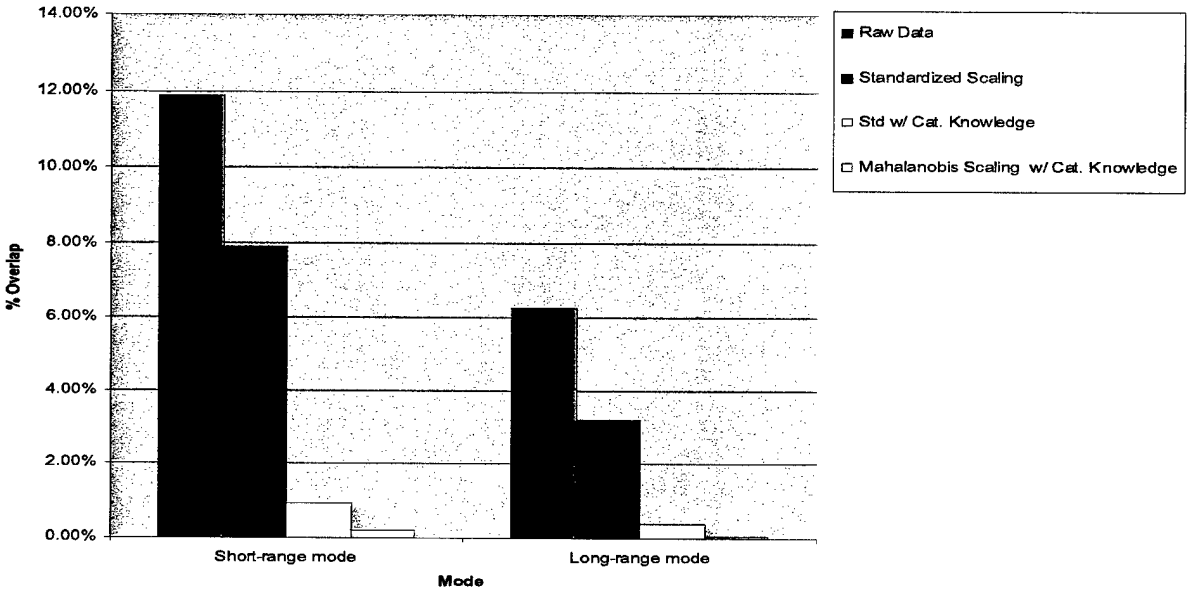

Fig. 7 — Percentage of overlap observed in >10,000 different-category distances with same-category distances for the three different measures used in the short- and long-range data, respectively.

3.4 Conclusion

There are two main implications of the above set of experiments. The first implication is that there is enough variation between different collects of the same emitter to make these collects statistically different. This was observed in the case examples and demonstrated quantitatively using the H-test. The second implication, however, is that this intraemitter variation is significantly less than the variation between collects of different emitters. Therefore, the data form used here is sufficient for emitter identification.

4. EID EXPERIMENTS

In this section, we discuss a set of experiments performed to analyze the structure of the data. The first set of experiments directly expanded our previous work [3] by using both conventional and IP parameters for EID, rather than just IP parameters alone. The second set of experiments compared the use of pulse averages over collects with individual pulse data for EID. The third set of experiments examined the use of scaling and Mahalanobis distance. The fourth set of experiments compared the different EID methods explored here. This included a comparison of Nearest Neighbor and SuperPHC with currently deployed systems.

4.1 Use of Both Conventional and IP Parameters for EID

This first set of experiments explored the use of both conventional and IP parameters for emitter identification. Three conventional parameters (PRI, RF, and PW) were used in EID, along with the 15 IP coefficients. This experiment is similar to our earlier work [3] except that here we also used conventional parameters, and pulse averages over collects were used in place of first identifying individual pulses then taking a majority vote.

4.1.1 Method

The Nearest Neighbor algorithm using Euclidean distance was used for EID. Data from both the short- and long-range mode collection were used. The short-range mode contained 42 different emitters in 163 collects, while the long-range mode contained 40 different emitters in 152 collects.

Experiments were performed independently for each of these data sets. A single pulse average was used to represent a single collect. A single trial was performed as follows: one collect was randomly selected to represent each emitter in a "training" set. The remaining collects were used to form the "test" set. Thus, for a trial using the long-range mode data set, there were 40 training collects and 112 testing collects. Likewise, for a trial using the short-range mode data set, there were 42 training sets and 121 testing sets. Next, we used Nearest Neighbor to identify the test set in both modes. The percentage of collects correctly identified resulted in an accuracy score for the single trial.

In all, 200 trials were performed for each of the data sets. The accuracy scores were averaged over these 200 trials. This process was repeated a total of 16 times, one for each possible permutation of use or nonuse of the four parameters (three conventional parameters with IP treated collectively as one super-parameter).

4.1.2 Results

Table 2 describes the identification accuracy of the Nearest Neighbor algorithm using the raw pulse averaged data as we used or ignored different parameters. Similar results were obtained with SuperPHC. The upper table shows long-range mode data set results while the lower table shows short-range mode results. Two-bit words describing the presence or absence of the four parameters (PW, IP, PRI, and RF) label the rows and columns. Higher performance was realized as more parameters were added. Where we used zero parameters for EID, we defined accuracy as that of pure guessing chance, which is respectively 1/40 and 1/42 for the long- and short-range mode data sets.

Table 2 — Percentages Correctly Identified in the Long- and Short-Range Modes

		PRI/RF (Long-range)			
		00	01	11	10
PW / IP	00	2.5%	22.1%	74.2%	46.1%
	01	71.9%	81.3%	92.4%	85.9%
	11	77.8%	85.3%	93.5%	89.2%
	10	14.8%	52.3%	82.6%	56.2%

		PRI/RF (Short-range)			
		00	01	11	10
PW / IP	00	2.4%	26.6%	46.6%	39.1%
	01	57.6%	60.8%	60.8%	57.6%
	11	57.7%	61.3%	60.6%	57.0%
	10	5.2%	30.7%	35.7%	15.9%

In the long-range mode, Table 2 tabulates the results of Nearest Neighbor using a plain Euclidean distance and pulse-averaged data. Each percentage displayed is an averaged accuracy over the 200 random perturbations of 40 training collects and the 112 blind test collects. As shown in the table, use of conventional parameters improved the accuracy by 21.6% compared to that of using the IP itself (93.5% vs 71.9%). Use of PRI alone enabled only 46.1% of the data to be identified correctly. Using RF alone, 22.1% of the data could be identified correctly. Pulse width was the least useful single parameter; it enabled us to correctly identify the emitter only 14.8% of the time.

The results in the short-range mode data are different from the results in the long-range mode data. The accuracy is lower in all the experiments and we see much less improvement with the addition of conventional parameters. For example, using IP alone we were able to identify the collects with 57.6% accuracy, increasing to only 60.8% when we added all of the conventional parameters. This contrasts with the long-range mode data, which improved from 71.9% accuracy to 93.5% accuracy when we added the conventional parameters.

For both data sets, the IP parameters appeared to contain the most information pertinent to EID, followed by the PRI information, and the RF information. The PW parameter was the least useful. This is easily seen in Figs. 8 and 9, which show clusters formed by individual parameters when taken from the 152 collects of the long-range mode data. Qualitatively, it can be seen in Fig. 8 that the PRI forms tighter clusters than does either the RF or the PW parameters.

4.2 Comparison of Single-Pulse and Pulse-Averaged EID

Next, we determined the EID accuracy attainable by using individual pulses and pulse averages. This was to verify the prediction from the H-test, above, that using individual pulses rather than pulse averages should not significantly improve EID accuracy.

4.2.1 Method

All of these experiments were performed with standardized data, rather than raw data as in the previous experiment. Both the Nearest Neighbor and SuperPHC algorithms were used. All three unconventional parameters and the IP parameters were used. We also followed the above method of randomly forming training and test sets from the data, and then identifying the test sets with the training sets. EID accuracies from a number of such trials were averaged to generate an overall EID accuracy. For the pulse-averaged data, 200 trials were performed, exactly as in the previous experiment. However, for the single-pulse data, the method had to be slightly varied due to the differing nature of single-pulse data.

In the single-pulse case, the training set was formed from all the pulses of the selected training collects of a trial. Then test collects were identified as follows: each pulse in a collect was first identified, then a majority vote was taken to choose the category of the entire collect. This method is the same as that followed in our previous work [3]. Only six trials were performed for the single-pulse case rather than 200. This is because the training and testing sets were each two or three orders of magnitude larger than those generated in the pulse-averaged categories, resulting in extremely CPU-intensive experiments.

Another variation that had to be made for the single-pulse experiments was the method of including PRI, which requires the use of multiple pulses to be meaningful. Here, we just used the single DTOA associated with each pulse in place of PRI. The first pulse of each collect was discarded because it had no meaningful DTOA.

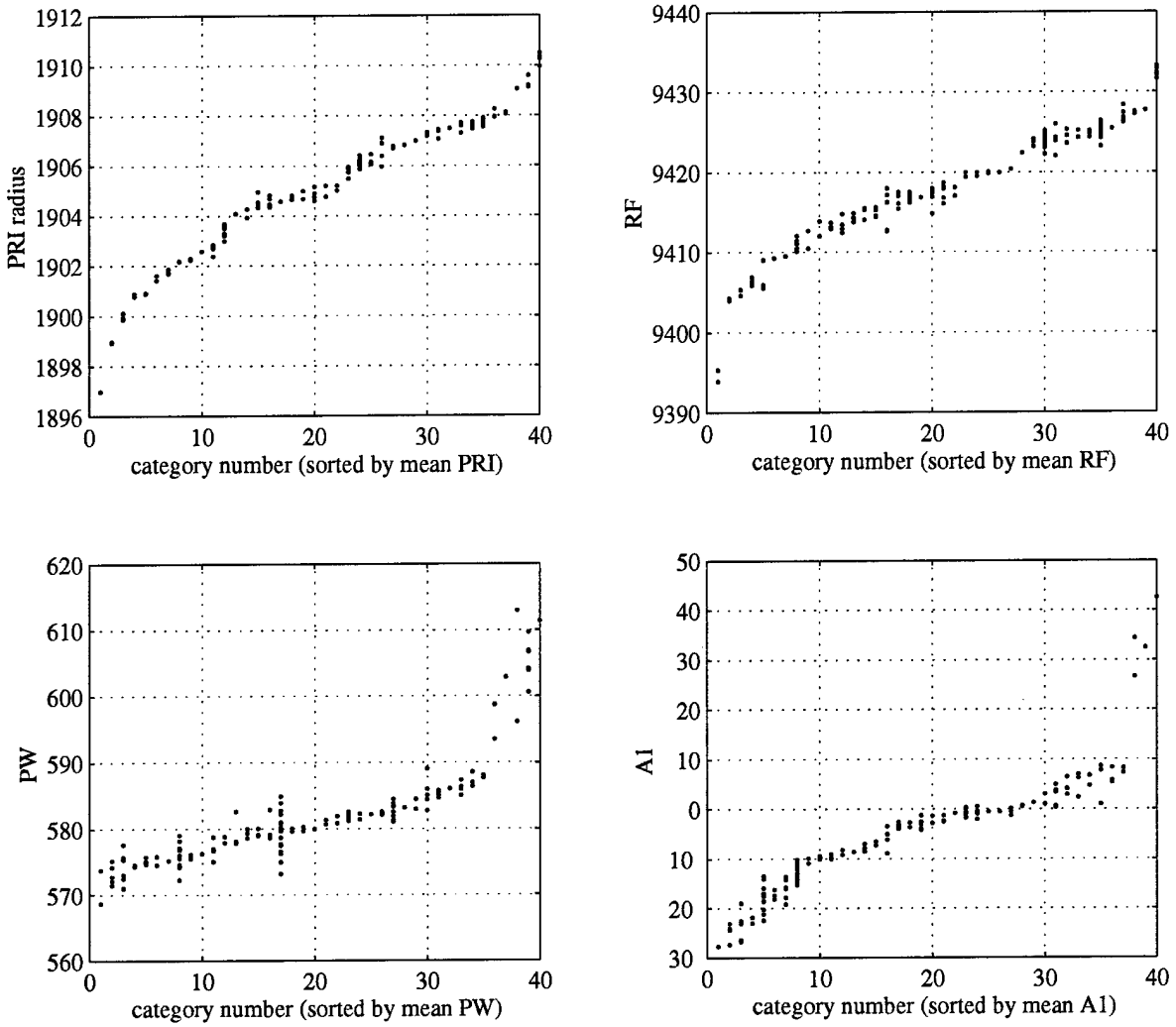

Fig. 8 — Distributions of four parameters (PRI, RF, PW, and IP parameter A1) for long-range mode data as sorted by the respective parameter means over the overall 40 categories. Thus, category 1 depicted here in each of the four panels does not necessarily represent category 1 in Appendix B. Also, only one IP parameter (A1) is represented in the figure.

4.2.2 Results

Table 3 shows a summary of the results. The EID accuracy overall, for both data subsets, improved from 72% [3] for unconventional parameters to 85% with the addition of the conventional parameters using SuperPHC when tested per pulse. The objective of the next experiment was to compare the EID accuracy attainable using pulse-by-pulse data vs pulse-averaged data. Note that these data are not directly comparable to Table 2 because here we used standardized data instead of the raw data.

Table 3 — EID Percent Accuracy Using Standardized Data

Mode	Nearest Neighbor		SuperPHC	
	Averaged	Per Pulse	Averaged	Per Pulse
Short-range mode	74.8%	77%	71%	73%
Long-range mode	94.0%	94%	90%	95%

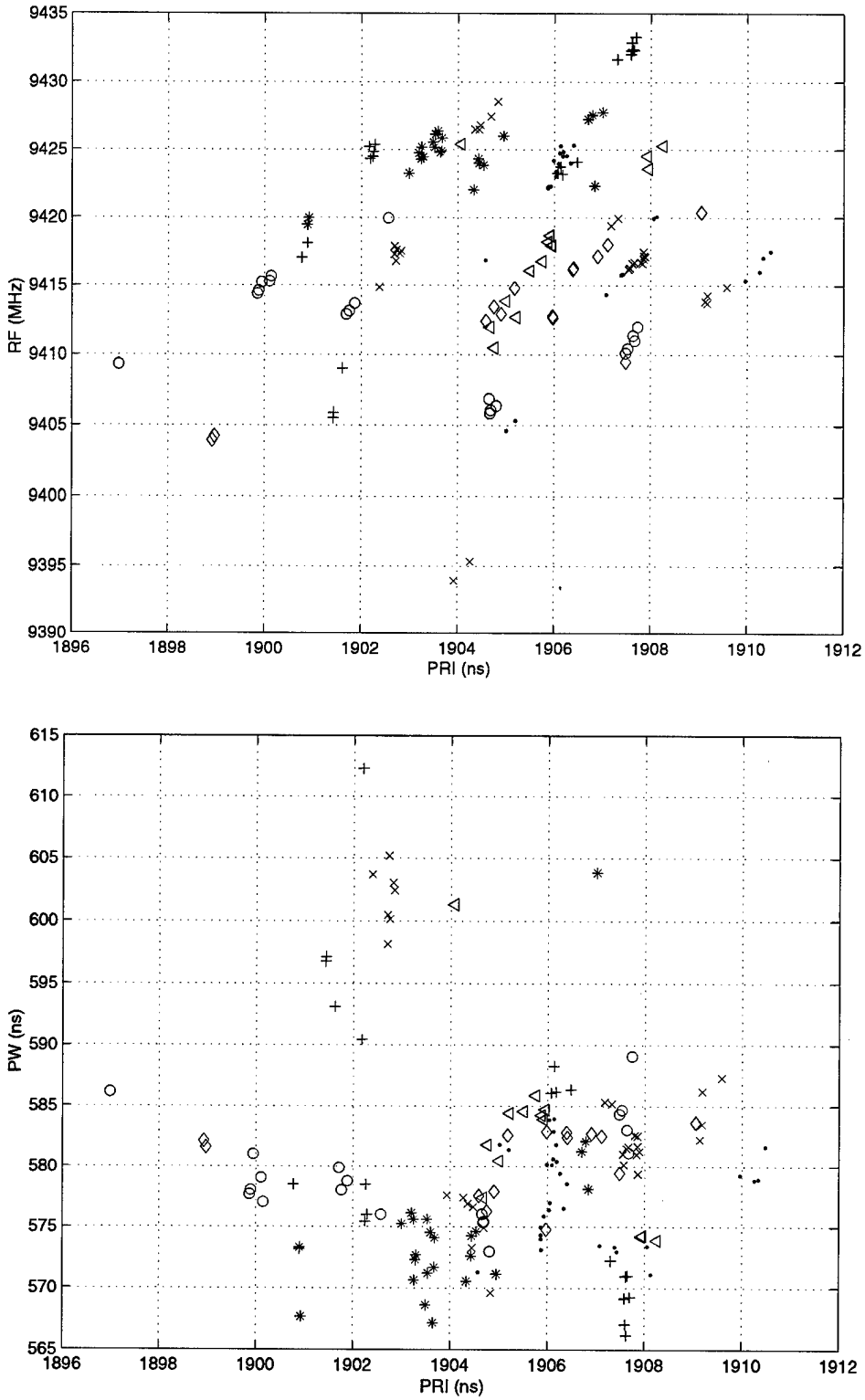

Fig. 9 — Plots of correlated and uncorrelated data from the 152 long-range mode collects. Top: Correlated RF vs PRI data. Each category is represented by a colored symbol. Individual collects of the same emitter have the same color and symbol. Bottom: PW vs PRI (uncorrelated data) of the 40 categories. Different colored symbols represent different categories.

When using the Nearest Neighbor algorithm, the accuracy using pulse-by-pulse EID was 2% higher than when using pulse averages. The accuracies were the same with the long-range mode data. This is a collective gain of 1%, which is not significant. The improvements were 2% when using SuperPHC. This is only a slight improvement, and if it were statistically significant, one would have to determine whether this slight increase in accuracy is worth the four to six orders of magnitude increase in computational complexity. The reader should note, though, that when using pulse-by-pulse EID, these high accuracies were obtained only after the majority vote. The actual EID accuracy was quite low, on the order of 60%. Overall, these results imply that the use of pulse averages for EID is essentially just as good as using pulse averages followed by majority vote. These results are consistent with the results of the H-test described above.

4.3 EID Accuracy Using Scaling and Mahalanobis Distance

Above we demonstrated that better EID accuracy can be obtained by using both conventional and IP parameters for EID, and that pulse averages are basically as effective for EID as single pulses. Here we further extend our work by using standardization and Mahalanobis distance rather than the simple Euclidean distance metric.

4.3.1 Method

These tests were performed by using both conventional and IP parameters in the same manner as above: one collect from each emitter was placed into the training set while the remaining collects formed the test set. The Nearest Neighbor EID algorithm was used. Pulse averages over entire collects were used. A total of 100 trials were performed for each case being tested. The EID accuracies obtained with the individual trials were averaged to form a final score. Tests were performed with raw data, standardized data, standardized data with categorical knowledge, data using the Mahalanobis distance, and data using Mahalanobis distance and categorical knowledge. In each trial, the statistics of the training set were used to form the scaling factors and the covariance matrix for the Mahalanobis distance.

Also, for four of these cases (all except Mahalanobis distance), overlaps between intracategory distances and interemitter distances were computed much in the same manner as in the previous section.

4.3.2 Results

Table 4 shows the EID results obtained from a test of Nearest Neighbor neural networks using different distance measures. Using raw data, we found that for the long-range mode data, Nearest Neighbor identified the data with 93.5% accuracy. However, with the raw short-range mode data, the accuracy was only 60.6%. The accuracy improved when we standardized the data by 14.2% for the short-range mode data, but by only 0.5% for the long-range mode data. Using data that were standardized with categorical knowledge, the accuracy was 93.3% for the short-range mode data and 98.5% for the long-range mode data. By standardizing the data, the Nearest Neighbor algorithm achieved a collective accuracy of 96.3% (by combining results from both subsets). When the data were identified with Nearest Neighbor using Mahalanobis distance instead of Euclidean with the covariance measured over the entire data set, the EID rates were both around 59%. When covariance matrixes were generated with categorical knowledge, the EID accuracy substantially improved to 98.5% for the short-range mode data and 100% for the long-range mode data. Clearly, categorical knowledge is necessary when using Mahalanobis distance.

4.4 Generalization of Scale Factors

Our practice here of using all the data to generate the covariance matrix or the scaling factors for our classifier is unrealizable in practice because we will not know categories before we identify the data, which was necessary in our previous experiment. Thus the above experiments do not address cases when new, previously unseen emitters are exposed. Here we explore how well the scale factors can be generalized using just a subset of the collects.

Table 4 – EID Percent Accuracy for Nearest Neighbor Tests
Using Averaged 18-Parameter Data

	Short-range mode	Long-range mode
Raw Data	60.6%	93.5%
Standardized	74.8%	94.0%
Standardized with Categorical Knowledge	93.3%	98.5%
Mahalanobis Distance	59.8%	58.6%
Mahalanobis Distance with Categorical Knowledge	98.5%	100.0%

4.4.1 Method

The same number of emitters was used to form training sets (40 from long-range mode data and 42 from short-range mode data). However, the number of emitters from the training sets used to form the scaling factors varied from 1 through 30. In each of these cases, 200 trials were performed in which test sets and training sets were randomly generated, and the appropriate number of training set collects were used to compute scaling factors. Pulse-averaged data and all available parameters were used for EID. The Nearest Neighbor algorithm was used. The goal was to observe how much EID accuracy degrades by using only a subset of the collects as categorical knowledge.

4.4.2 Results

Figure 10 shows the EID accuracy as we increase the number of categories used to calculate the covariance matrix for our Mahalanobis distance network. Accuracy increases dramatically up to 11 categories, then much less up to 30 categories. Each point on the graph is an average accuracy over the 200 trials. The results show that the scale factors do in fact generalize well but not well enough to obtain the high accuracy seen in the previous experiment (Table 4). Such a high accuracy required that all the emitters be used to form scaling factors.

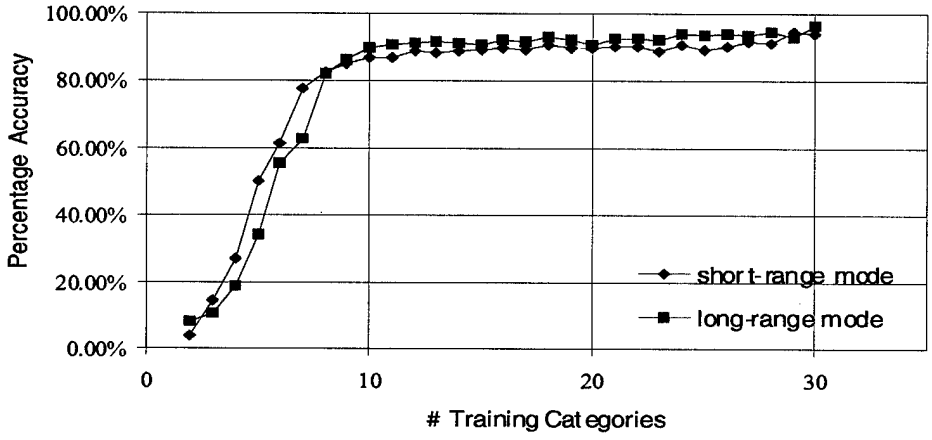

Fig. 10 — Percentage accuracy vs number of categories used to calculate the covariance matrix

Let us examine the correlation matrix in further detail. Table 5 lists some of the correlation coefficients. These were taken from the Mahalanobis distance with categorical knowledge matrix σ_{Li}, computed in Section 2.3.5. The strong positive correlation between PRI and RF within categories is of interest; this is depicted in Fig. 9. For comparison, PRI vs PW is also given to illustrate uncorrelated parameters.

Note that in these results, the covariance matrix is ill-conditioned with fewer than 10 training categories, which also results in poor EID accuracy. Generally, one requires for an $n \times n$ covariance matrix, $n(n+1)/2$ points of data, and here, fewer than that are available. As shown in Fig. 10, 94% accuracy can be achieved in both data modes as up to 30 out of 40 categories become available for training.

Table 5 – Correlation Between Parameters in the Long-Range Mode Data

	PRI	RF	PW	A1	A2
PRI	1.0000	X	X	X	X
RF	0.8450	1.0000	X	X	X
PW	0.0755	0.0922	1.0000	X	X
A1	-0.3632	-0.4564	0.3227	1.0000	X
A2	-0.2622	-0.3036	-0.1545	0.2663	1.0000

4.5 Comparison of Different EID Algorithms

Next we compare EID accuracies obtained above with those obtainable by deployed systems. This is done to examine how current systems can be improved. We also compare Nearest Neighbor with SuperPHC in terms of EID speed. This is done to examine whether the conclusions in our previous work [3] regarding SuperPHC remain valid in this study.

4.5.1 Method

The two algorithms used in baseline systems are the Matching Algorithm and the Normalized Euclidean Metric (see Section 2). These two algorithms use only IP information. To test their effectiveness, we compared them to the SuperPHC and Nearest Neighbor algorithms. Three variants of these second two algorithms were used: standardized data, data standardized with categorical knowledge, and Mahalanobis distance computed with categorical knowledge. Rather than randomly selecting training and testing sets as in the above experiments, these experiments were performed using the jackknifing strategy described in Section 2. The experiments were performed separately with the two modes. The final EID accuracy was obtained by averaging the results of two separate sizes. The comparison of EID speed between Nearest Neighbor and SuperPHC was performed on the Sun Workstation. The tests for the study consisted of 150 pulse-averaged samples for training and 1 through 150 pulse-averaged samples for testing. SuperPHC required training; therefore, the training time was measured. However, to obtain a fuller appreciation of the relative EID time for the two methods on a larger dataset, both single-pulse representation and pulse-averaged data were used.

4.5.2 Results

Table 6 presents the final EID results obtained. We see that by using IP alone, the Matching Algorithm and the normalized Euclidean metric can identify the emitters with 84.1% and 85.8% accuracy, respectively. The distance measures we used in most of our experiment yield comparable levels of accuracy. The accuracy improves to 91.6% by using Mahalanobis distance with categorical knowledge instead of Euclidean distance. However, the accuracy improves substantially if we use both conventional and IP parameters—from 91.6% to 99.3%. This implies that deployed systems can improve by using the methods developed here. We also note that Nearest Neighbor performs slightly better than SuperPHC, but by only a small amount (99.3% vs 98.6%).

Above we used both the Nearest Neighbor and SuperPHC algorithms for EID. As was found in a previous report [3], SuperPHC has very similar EID accuracy to Nearest Neighbor but does not exceed it. The above results show the same pattern. Also, in the previous work it was shown that SuperPHC was signifi-

cantly faster than Nearest Neighbor in testing, hence is ideal for certain applications. SuperPHC was also significantly faster in EID when the training sets were more complex. Note however, that the scaleup in time was linear for the 300 pulse-averaged representation.

Table 6 – Percent Identified Correctly out of 296 Test Files

		IP Alone	All Parameters
Conventional Algorithms	**Matching Algorithm**	84.1%	-
	Normalized Euclidean Metric	85.8%	-
SuperPHC	Standardized	81.1%	92.9%
	Standardized with Categorical Knowledge	87.1%	98.0%
	Mahalanobis Distance with Categorical Knowledge	89.5%	98.6%
Nearest Neighbor	Standardized	84.5%	93.9%
	Standardized with Categorical Knowledge	85.1%	98.3%
	Mahalanobis Distance with Categorical Knowledge	91.6%	99.3%

Table 7 demonstrates this effect by presenting the times for training and testing for a set of 100 to 10,000 pulses. The training time grew linearly for a small pulse set (fewer than 1000 pulses). However, for sets larger than 1000 pulses, the training time for SuperPHC and the EID time using the Nearest Neighbor algorithm was at least an order of magnitude slower (more than 100 s vs less than 10 s for 1000 pulses). Note that the EID time grew linearly for SuperPHC since training on larger collects could be done effectively offline.

Table 7 — EID Time for Nearest Neighbor and SuperPHC on a Sun Ultra Workstation

	# Pulses	Training Time (s)	Test Time (s)
Nearest Neighbor	100	N/A	2
	1000	N/A	7
	10,000	N/A	702
SuperPHC	100	3.5	1
	1000	7	2
	10,000	180	5

5. DISCUSSION

In the current study of emitter identification using field data from 42 Raytheon RX41 navigational radars, EID accuracy exceeded 90% using IP alone when an appropriate scaling factor was used. Higher EID rates were observed with long-range mode over the short-range mode data. With the addition of three conventional parameters, an increase in EID accuracy was noted in both pulse-averaged and pulse-by-pulse experiments. Overall, the accuracies quoted in the previous work [3] can be improved by either using scaling factors or by adding conventional parameters. The latter validates the prediction made in Ref. 4. By

using both appropriate scaling without categorical knowledge and all (both conventional and IP) parameters, the overall EID accuracy exceeded 95%. A collective EID accuracy of greater than 99% was obtained using Mahalanobis distance with categorical knowledge.

Nearest Neighbor and SuperPHC gave comparable EID rates. Increased EID rates were observed for the long-range mode collects, but not for short-range mode collects, when using both algorithms. There are several explanations for this fact. First of all, there are more emitters/categories in the short-range mode collects. More categories clearly imply an increased chance of confusion during EID. Moreover, since pulses were sampled at 6.25 ns intervals, fewer raw pulses (32 vs 64 for long-range mode collects) provide the information for a short-range mode [3].

It has been hypothesized that the selection of a short-range mode for sampling may cut off useful discriminating information located in the latter part of the pulse. The lower performances of the algorithms in the short-range mode hash bin could be a result of this effect. However, under certain scaling conditions, no measurable difference was observed (Table 6).

5.1 Conclusions

The following paragraphs focus on the most important points that can be drawn from the work described above.

5.1.1 Conventional Parameters Improve EID Accuracy

Higher EID rates were observed throughout with the addition of conventional parameters (Tables 1, 4, and 6). This happened whether or not scaling factors were used. This validates the prediction made in Ref. 5, which predicted that accuracies approaching 99% should be achievable when all parameters were used. Without scaling factors, the addition of conventional parameters improved EID accuracy from 72% to 93.5% on the long-range mode data and from 57.6% to 60.6% on the short-range mode data (Table 2). When jackknifing was used and the parameters were standardized without categorical knowledge, a collective accuracy (short and long-range mode) rose from 84.5% with IP parameters alone to 93.9% with all parameters (Table 6). Similar improvements were obtained by using categorical knowledge either with standardization or Mahalanobis distance. These improvements were independent of whether Nearest Neighbor or SuperPHC was used for EID.

5.1.2 Effects of Pulse Averaging

A detailed examination of training with averaged and single pulse representation revealed effectively no difference in accuracy gained. This was consistent with the results of the H-test. The Kruskal-Wallis H-test showed that there is not only random variation between pulses within a collect, but that there is also a second random process that changed the distributions of the pulse parameters between different collects of the same emitter. This second random process seemed to change the pulses to a greater extent than accountable by the inherent intracollect random process variation. Therefore, using individual pulses within a collect did not necessarily improve EID accuracy. This was verified experimentally, where keeping track of the distribution of collects has negligible effects on the accuracy.

One could argue from the results shown in Table 3 that single-pulse representations performed slightly better than did pulse averages. This increase was on the order of 1% or 2%, however, and thus appears to be relatively insignificant. Further studies would have to be performed to determine if this increase is indeed reproducible and reliable. However, even if this increase is statistically significant, the huge increase in required CPU time and the huge increase in neural network size may not justify this slight increase in accuracy: comparing pulse-averaged data are approximately 15,000 times less computationally demanding and 800 times less memory intensive.

5.1.3 Scaling With Categorical Knowledge is Essential for Highly Accurate (95% to 99%) EID

Standardizing the data certainly improves EID accuracy, as discussed above. However, highly accurate EID requires the use of categorical knowledge with either scaling or Mahalanobis distance. Qualitatively, this makes sense because the use of categorical knowledge makes the scaling factors more sensitive to the statistics of individual collects rather than to the displacements between collects. Thus the "dividing lines" set up in parameter space to discriminate between different categories are positioned in a way that is sensitive to the cluster's shape. However, it is necessary to determine whether the scaling factors are generalizable. We demonstrated that the scaling factors are indeed generalizable but only up to a point. Unfortunately in many real mission scenarios there will be many collects that are previously unseen. The results here do not guarantee that such emitters will be identified with great accuracy. If the mission scenario is such that all emitters have been seen previously, then we know that highly accurate EID can be performed.

The EID schemes used here may be difficult to apply to other EID data. The PRIs will not always be a two-position stagger as was the case of the emitters from this experiment. It may be much more complex, having many more PRI legs and jitter. However, if PRI is used in some other scheme, it could be made to give a confidence factor to be combined with EID based on other parameters. Although PRI is correlated with other parameters, it has also been shown that by proper scaling, a distance measure that does not take this into account can also achieve similar accuracy. One example is a 1.5% difference between the Mahalanobis distance with categorical knowledge and the standardization with categorical knowledge (Table 4).

5.1.4 Both Algorithms Have Comparable EID Accuracy

SuperPHC yields similar EID accuracy to Nearest Neighbor throughout our experiments (Tables 3 and 6). Essentially, SuperPHC shows slightly less accuracy than Nearest Neighbor over all the different permutations of experiments performed. The drawback of SuperPHC is that it requires training. This is especially true when the set consists of more than 1000 training vectors. However, we have seen that pulse averaging yields similar results to those obtained in pulse-by-pulse training experiments (Table 3). Thus, in the pulse-averaged case with less than 500 training vectors, Nearest Neighbor may be the algorithm of choice. This is offset by the fact that as there are more and more training examples, Nearest Neighbor grows linearly in EID time, but SuperPHC, since it divides the data into hierarchies, requires less time in EID. The distinct advantage given by SuperPHC, as shown in our previous work [3], is that for very large databases it outperforms Nearest Neighbor in terms of EID speed. This is observed again in Table 7.

5.1.5 Comparison with Currently Used EID Schemes

Using intrapulse parameters alone, 90% EID accuracy was noted with Mahalanobis distance (Table 6). This is at least 5% better than the Matching Algorithm currently used in the field [10]. Other EID schemes using IP alone yield rates around 85%. By including all parameters, we routinely observed an accuracy greater than 95%. Pulse-averaging reduced the training time without loss of EID accuracy. Furthermore, an entire collect of 896 pulses can be analyzed in less than one second. Since training time is a critical factor for a tactical ESM system, then Nearest Neighbor may be the algorithm of choice. If, however, training can be performed offline, then SuperPHC would be beneficial with rapid EID of test pulses using simple processors.

5.2 Summary of Practical Implications

The above results can be distilled into the following practical implications:

1) Use of conventional and intrapulse parameter information improves the ID performance of matching algorithms. Also, pulse averages rather than individual pulses should be used in library development.

2) ID accuracies can be improved by appropriate standardization of data.
3) If a specific mission is such that all or most of the emitters are previously unseen, categorical knowledge can be used with standardization or Mahalanobis distance to significantly improve EID accuracy. Categorical knowledge represents library entries of previous intercepts. Thus, if emitters have been previously intercepted and entered into the library, then accuracy of EID of new intercepts is high. If emitters have not been previously intercepted, then improper association with existing entries is relatively low (i.e., the novelty detection is relatively good).
4) Greater than 95% EID accuracy can be achieved by appropriate standardization or by using Mahalanobis distance with categorical knowledge can be had with either SuperPHC or Nearest Neighbor. Thus both algorithms can be deployed. SuperPHC will be of choice in a dense emitter environment if training can be done offline.

5.3 Research Directions and Future Work

One interesting research possibility is the study of PRI angles for EID. All of the emitters collected here were from one model, and they all had identical PRI angles. This is most likely a design feature and it is possible that a different manufacturer of emitters would choose different PRI angles for their emitters. This could lead to a quick method for identifying a specific emitter, which may be valuable in certain situations.

Improvements in EID accuracy in the single pulse case would be more difficult to achieve, if all pulses were to be included; in many cases, the IP vectors were not separable or were too corrupted by noise. To address this problem, adaptive feature extraction methods can be applied to the raw I&Q postdetector signal. It is believed that this research will reveal better features and pulse representations to create a more effective IP vector. This will allow finer discrimination while being more robust against noise.

Our experiments imply the following in an automatic EID system: pulse averaging is an efficient method for rapid EID, and scaling the data improves EID accuracy. Both classifier algorithms outperformed those currently used. Nearest Neighbor is a better choice if parallel computing is available because it uses downloaded data to directly update the network. SuperPHC requires training but results in a faster classifier than does Nearest Neighbor, especially on a serial processor.

REFERENCES

1. R.G. Wiley, *Electronic Intelligence: The Analysis of Radar Signals*, 2nd edition (Artech House, Boston, 1993).

2. J.C. Sciortino, S. Yang, M.J. Thompson, G.L. Barrows, D.A. Stenger, and V.C. Kowtha, "Highly Accurate Autonomous ESM Surveillance Technology for System Applications," Proceedings of the 44th Annual Joint Electronic Warfare Conference, April 28, 1999, Naval Air Station, Pensacola, FL.

3. G.L Barrows, J.C. Sciortino, V.C. Kowtha, and D.A. Stenger, "Specific Emitter Identification Using Massively Parallel Implementations of Neural Networks," Naval Research Laboratory Report NRL/FR/5720—96-9830, Sept. 1996.

4. J.C. Sciortino, "Adaptive, Self-Directed ESM Surveillance Technology for System Applications," Proceedings of the 41st Annual Joint Electronic Warfare Conference, May 13-15, 1996, United States Naval Postgraduate School, Monterey, CA, 1996.

5. J.C. Sciortino, G.L. Barrows, V.C. Kowtha, and D.A. Stenger, "Towards Highly Accurate Emitter Identification," Proceedings of the National Fire Control Symposium, August 4-7, 1997, Vol. II, pp. 469-488.

6. J.C. Sciortino, Jr., "Autonomous ESM Systems," *Naval Engineers Journal* **109**, 73-83, 1997.

7. G.L. Barrows and J.C. Sciortino, "A Mutual Information Measure for Feature Selection with Application to Pulse EID," Proceedings of IEEE Intl. Symposium on Time-Frequency and Time-Scale Analysis, June 18-21, 1996, Paris, France.

8. K.S. Younis, "Weighted Mahalanobis Distance for Hyper-Ellipsoidal Clustering," Air Force Inst. of Technology Pub. # 96D-22, 1996.

9. A. Teolis, "L-MISPE/SMDA Coefficient Compatibility Testing," AIMS TR-97-03, AIMS, Inc., March 1997.

10. R. Mentle and A. Teolis, "LMISPE Analysis," NRL Monthly Report, AIMS, Inc., Sept. 1996.

11. J. Ambros-Ingerson, R. Granger, and G. Lynch, "Simulation of Paleocortex Performs Hierarchical Clustering," *Science* **247**,1344-1348, 1990.

12. R. Granger, J. Whitson, and E. Whepley, "SuperPHC," proprietary computer program written in C by Thuris Corporation, Inc., Irvine, CA, 1992.

13. R.G. Wiley, "Final Musketeer-Dixie II Data Analysis," Hanover, MD, 1993.

14. L. Sachs, *Applied Statistics: A Handbook of Techniques*, 2nd ed. (Springer-Verlag, New York, 1984).

15. K. Fukunaga, "Introduction to Statistical Pattern Recognition," in *Computer Science and Scientific Computing*, W. Rheinbolt, ed. (Academic Press, Boston, 1990).

Appendix A

SHORT-RANGE MODE COLLECT DATA

Category	Collect#	RF	Date	Time	Location	ID
01	590	9426.6	12	1040	PP	A1
02	589	9415.5	12	1026	PP	A3
03	284	9419.3	07	1044	MCT	A8
03	285	9419.3	07	1044	MCT	A8
04	320	9408.7	08	0820	SLA	B3
04	621	9406.0	12	1606	SLA	B3
05	535	9420.6	11	1047	PP	B8
06	010	9423.8	01	1030	SLA	C1
06	011	9423.8	01	1030	SLA	C1
06	178	9423.4	04	1604	SLA	C1
06	228	9423.5	06	0934	SLA	C1
06	417	9422.9	09	1508	SLA	C1
06	502	9422.6	10	1908	SLA	C1
06	588	9420.7	12	1010	SLA	C1
07	271	9421.8	07	0819	SLA	C3
07	622	9419.7	12	1618	SLA	C3
08	229	9435.4	06	0949	MCT	C8
08	230	9435.4	06	0949	MCT	C8
09	023	9412.3	01	1348	SLA	D1
09	013	9412.9	01	1057	SLA	D1
09	014	9412.9	01	1057	SLA	D1
09	068	9412.1	02	1142	SLA	D1
09	069	9412.1	02	1142	SLA	D1
09	139	9412.8	03	1642	SLA	D1
09	185	9413.4	05	0843	SLA	D1
09	268	9413.8	06	1731	SLA	D1
09	315	9412.3	07	1636	SLA	D1
09	323	9413.4	08	0847	SLA	D1
09	427	9412.3	09	1648	SLA	D1
09	484	9411.1	10	1524	SLA	D1
09	549	9412.4	11	1405	SLA	D1
10	312	9420.3	07	1606	SLA	D8
10	563	9420.1	11	1532	SLA	D8
11	338	9427.6	08	1034	PP	E5
12	020	9426.3	01	1322	SLA	F1
12	021	9426.3	01	1322	SLA	F1
12	096	9424.8	02	1631	SLA	F1
12	097	9424.8	02	1631	SLA	F1
12	198	9424.9	05	1053	SLA	F1
12	426	9425.3	09	1635	SLA	F1
12	627	9423.8	12	1702	SLA	F1
12	628	9423.8	12	1702	SLA	F1
13	391	9410.9	09	0841	SLA	F3

Category	Collect #	RF	Date	Time	Location	ID
13	599	9410.1	12	1331	SLA	F3
13	600	9410.1	12	1331	SLA	F3
13	611	9409.3	12	1454	SLA	F3
14	339	9427.2	08	1054	SB	F8
14	340	9427.2	08	1054	SB	F8
14	365	9427.5	08	1603	SB	F8
15	376	9413.3	08	1850	SLA	G3
15	580	9415.7	12	0853	SLA	G3
15	620	9413.6	12	1555	SLA	G3
16	434	9425.4	10	0906	MCT	G8
16	435	9425.4	10	0906	MCT	G8
16	505	9424.1	10	1950	MCT	G8
16	506	9424.1	10	1950	MCT	G8
17	543	9428.9	11	1148	SLA	H8
18	151	9419.7	04	0955	PP	I1
18	152	9419.7	04	0955	PP	I1
18	200	9419.1	05	1127	PP	I1
18	348	9418.2	08	1307	SLA	I1
18	429	9420.1	10	0817	SLA	I1
18	430	9420.1	10	0817	SLA	I1
18	504	9417.6	10	1936	SLA	I1
19	457	9422.4	10	1135	SB	J8
19	472	9423.1	10	1408	SB	J8
19	473	9423.1	10	1408	SB	J8
20	597	9430.5	12	1147	PP	K3
21	049	9426.1	02	0908	MCB	L1
21	050	9426.1	02	0908	MCB	L1
21	113	9428.6	03	1102	MCB	L1
21	114	9428.6	03	1102	MCB	L1
21	153	9428.6	04	1014	MCB	L1
21	154	9428.6	04	1014	MCB	L1
21	155	9428.6	04	1014	MCB	L1
21	255	9427.5	06	1457	MCT	L1
21	256	9427.5	06	1457	MCT	L1
21	343	9425.9	08	1125	PP	L1
21	541	9426.9	11	1134	PP	L1
21	557	9427.6	11	1511	PP	L1
22	494	9427.0	10	1636	SLA	L3
23	072	9418.9	02	1305	MCT	M1
23	071	9418.9	02	1305	MCT	M1
23	041	9418.0	01	1619	MCT	M1
23	091	9418.6	02	1558	MCT	M1
23	092	9418.6	02	1558	MCT	M1
23	128	9419.2	03	1456	MCT	M1
23	129	9419.2	03	1456	MCT	M1
23	145	9419.5	04	0846	MCT	M1
23	146	9419.5	04	0846	MCT	M1
23	626	9417.2	12	1649	SLA	M1
24	286	9422.9	07	1059	PP	M8

Category	Collect #	RF	Date	Time	Location	ID
24	353	9420.9	08	1405	PP	M8
24	412	9423.7	09	1422	PP	M8
25	157	9431.7	04	1047	PP	N1
25	158	9431.7	04	1047	PP	N1
25	159	9431.7	04	1047	PP	N1
25	176	9432.9	04	1546	PP	N1
25	177	9432.9	04	1546	PP	N1
25	308	9427.5	07	1534	PP	N1
25	309	9427.5	07	1534	PP	N1
25	530	9429.2	11	1011	PP	N1
26	043	9426.3	02	0807	SLA	O1
26	181	9425.2	04	1654	SLA	O1
26	205	9424.6	05	1427	SLA	O1
26	511	9423.9	11	0814	SLA	O1
27	371	9427.1	08	1819	SLA	O3
27	372	9427.1	08	1819	SLA	O3
27	478	9425.8	10	1439	SLA	O3
28	419	9397.4	09	1538	PP	P3
29	062	9416.4	02	1053	SLA	Q1
29	063	9416.4	02	1053	SLA	Q1
29	215	9416.6	05	1613	SLA	Q1
29	387	9415.5	08	2002	SLA	Q1
30	316	9412.5	07	1651	SLA	Q3
30	439	9414.6	10	1009	SLA	Q3
30	608	9412.7	12	1429	SLA	Q3
31	593	9408.8	12	1114	MCT	R1
32	591	9417.9	12	1059	PP	S1
32	592	9417.9	12	1059	PP	S1
33	118	9416.0	03	1311	SLA	T1
33	119	9416.0	03	1311	SLA	T1
33	173	9416.0	04	1457	SLA	T1
33	266	9416.0	06	1659	SLA	T1
33	565	9414.2	11	1556	SLA	T1
34	100	9414.7	03	0826	SLA	W1
34	101	9414.7	03	0826	SLA	W1
34	221	9414.1	06	0814	SLA	W1
34	517	9412.4	11	0901	SLA	W1
35	324	9423.5	08	0903	SLA	W3
35	507	9421.6	10	2003	SLA	W3
36	099	9415.4	03	0809	SLA	X1
36	240	9414.8	06	1124	SLA	X1
36	241	9414.8	06	1124	SLA	X1
36	385	9414.1	08	1935	SLA	X1
36	428	9414.9	09	1702	SLA	X1
37	133	9420.9	03	1534	PP	Y1
37	134	9420.9	03	1534	PP	Y1
37	197	9421.4	05	1038	PP	Y1
37	298	9419.8	07	1358	PP	Y1
37	437	9421.0	010	0940	PP	Y1

Category	Collect #	RF	Date	Time	Location	ID
38	143	9419.5	04	0829	SLA	Z1
38	144	9419.5	04	0829	SLA	Z1
38	220	9418.3	05	1713	SLA	Z1
38	227	9418.6	06	0918	SLA	Z1
38	528	9416.9	11	0959	SLA	Z1
38	529	9416.9	11	0959	SLA	Z1
39	106	9426.4	03	0933	SLA	J1
39	107	9426.4	03	0933	SLA	J1
39	164	9429.3	04	1301	SLA	J1
39	165	9429.3	04	1301	SLA	J1
39	246	9427.4	06	1215	SLA	J1
39	389	9429.1	09	0827	SLA	J1
39	390	9429.1	09	0827	SLA	J1
40	109	9417.3	03	1007	PP	P1
40	234	9418.7	06	1051	PP	P1
40	235	9418.7	06	1051	PP	P1
40	349	9416.6	08	1320	SLA	P1
40	495	9413.9	10	1651	SLA	P1
40	598	9415	12	1202	SLA	P1
41	370	9408.2	08	1704	SLA	Y3
42	366	9411.4	08	1620	SLA	Z3

Appendix B

LONG-RANGE MODE COLLECT DATA

Category	Collect#	RF	Date	Time	Location	ID
01	002	9425.6	01	0945	SLA	A2
01	016	9422.1	01	1119	SLA	A2
01	037	9423.3	01	1559	SLA	A2
01	038	9423.3	01	1559	SLA	A2
01	039	9423.3	01	1559	SLA	A2
01	040	9423.3	01	1559	SLA	A2
01	102	9425.1	03	0843	SLA	A2
01	123	9423.6	03	1402	SLA	A2
01	124	9423.6	03	1402	SLA	A2
01	138	9424.4	03	1627	SLA	A2
01	174	9424.8	04	1512	SLA	A2
01	252	9423.4	06	1422	SLA	A2
01	270	9424.4	06	1746	SLA	A2
01	296	9422.7	07	1325	SLA	A2
01	342	9423.3	08	1112	PP	A2
01	421	9423.9	09	1553	PP	A2
01	470	9424.1	10	1335	PP	A2
01	534	9423.3	11	1036	PP	A2
02	351	9414.1	08	1336	PP	A4
02	546	9414.4	11	1320	PP	A4
02	602	9414.0	12	1343	PP	A4
03	295	9415.7	07	1308	MCT	A9
03	375	9415.2	08	1834	MCT	A9
03	384	9414.4	08	1919	MCT	A9
03	425	9416.5	09	1620	MCT	A9
04	503	9403.8	10	1922	SLA	B4
04	576	9404.7	12	0820	SLA	B4
05	491	9416.0	10	1617	PP	B9
06	249	9420.0	06	1350	SLA	C2
06	532	9418.7	11	1024	SLA	C2
07	194	9418.5	05	1007	SLA	C4
07	195	9418.5	05	1007	SLA	C4
07	317	9415.8	07	1705	SLA	C4
07	499	9416.7	10	1837	SLA	C4
07	630	9416.1	12	1713	SLA	C4
08	189	9433.0	05	0916	MCT	C9
08	190	9433.0	05	0916	MCT	C9
08	204	9431.6	05	1410	MCT	C9
08	223	9431.8	06	0845	MCT	C9
08	224	9431.8	06	0845	MCT	C9
08	244	9432.0	06	1200	MCT	C9
08	245	9432.0	06	1200	MCT	C9
09	307	9416.3	07	1516	SLA	D9

Category	Collect #	RF	Date	Time	Location	ID
09	537	9417.2	11	1058	SLA	D9
10	253	9424.8	06	1441	PP	E9
10	254	9424.8	06	1441	PP	E9
10	291	9424.1	07	1131	PP	E9
11	110	9423.0	03	1023	SLA	F2
11	267	9423.2	06	1715	SLA	F2
11	413	9422.3	09	1439	SLA	F2
11	575	9423.0	12	0806	SLA	F2
12	418	9407.7	09	1522	SLA	F4
12	616	9406.0	12	1519	SLA	F4
12	618	9405.6	12	1530	SLA	F4
13	328	9424.6	08	1003	SB	F9
13	362	9423.6	08	1515	SB	F9
14	272	9411.5	07	0835	SLA	G4
14	497	9409.2	10	1703	SLA	G4
15	501	9421.0	10	1853	MCT	G9
16	395	9426.4	09	0912	SLA	H9
16	516	9426.8	11	0848	SLA	H9
17	042	9422.0	11	1637	SLA	J2
17	201	9423.7	05	1142	SLA	J2
17	214	9424.4	05	1559	SLA	J2
17	259	9424.3	06	1547	SLA	J2
17	431	9424.8	10	0835	SLA	J2
18	450	9419.1	10	1100	SB	J9
18	483	9418.1	10	1511	SB	J9
19	587	9426.1	12	0955	PP	K4
20	073	9424.9	02	1325	MCB	L2
20	074	9424.9	02	1325	MCB	L2
20	075	9424.9	02	1325	MCB	L2
20	104	9426.3	03	0916	MCB	L2
20	105	9426.3	03	0916	MCB	L2
20	135	9427.0	03	1552	MCB	L2
20	147	9425.2	04	0903	MCB	L2
20	148	9425.2	04	0903	MCB	L2
20	199	9426.1	05	1111	MCT	L2
20	213	9425.6	05	1543	MCT	L2
20	264	9425.8	06	1640	MCT	L2
20	265	9425.8	06	1640	MCT	L2
20	551	9422.0	11	1419	PP	L2
21	273	9424.4	07	0851	SLA	L4
22	045	9417.0	02	0828	MCT	M2
22	046	9417.0	02	0828	MCT	M2
22	047	9417.0	02	0828	MCT	M2
22	087	9416.6	02	1518	MCT	M2
22	088	9416.6	02	1518	MCT	M2
22	089	9416.6	02	1518	MCT	M2
22	103	9417.3	03	0859	MCT	M2
22	115	9416.4	03	1120	MCT	M2
22	156	9417.5	04	1029	MCT	M2

Category	Collect #	RF	Date	Time	Location	ID
22	579	9415.5	12	0834	SLA	M2
23	257	9420.0	06	1512	PP	M9
23	398	9418.4	09	0926	PP	M9
24	058	9427.4	02	1021	PP	N2
24	131	9429.5	03	1513	PP	N2
24	132	9429.5	03	1513	PP	N2
24	359	9425.9	08	1447	PP	N2
24	480	9425.0	10	1454	PP	N2
25	051	9418.1	02	0926	PP	P2
25	052	9418.1	02	0926	PP	P2
25	054	9418.1	02	0926	PP	P2
25	175	9417.7	04	1528	PP	P2
25	211	9417.8	05	1512	PP	P2
25	260	9417.8	06	1604	PP	P2
25	498	9413.5	10	1718	SLA	P2
26	444	9394.3	10	1042	PP	P4
26	604	9393.9	12	1355	PP	P4
27	048	9413.9	02	0848	SLA	Q2
27	120	9414.4	03	1328	SLA	Q2
27	121	9414.4	03	1328	SLA	Q2
27	388	9413.7	09	0811	SLA	Q2
28	322	9412.9	08	0833	SLA	Q4
28	488	9410.6	10	1603	SLA	Q4
29	064	9406.3	02	1110	SLA	R2
29	065	9406.3	02	1110	SLA	R2
29	219	9406.5	05	1659	SLA	R2
29	614	9406.6	12	1506	MCT	R2
30	369	9408.2	08	1648	SLA	R4
31	066	9414.4	02	1126	PP	S2
31	067	9414.4	02	1126	PP	S2
31	090	9415.2	02	1540	PP	S2
31	345	9414.7	08	1139	PP	S2
31	540	9414.0	11	1119	PP	S2
32	079	9413.1	02	1359	SLA	T2
32	203	9413.5	05	1355	SLA	T2
32	386	9411.7	08	1949	SLA	T2
33	548	9418.9	11	1351	PP	T4
34	137	9411.0	03	1609	SLA	U2
34	160	9412.0	04	1120	SLA	U2
34	217	9411.0	05	1644	SLA	U2
34	574	9410.6	11	1648	SLA	U2
34	625	9409.9	12	1639	SLA	U2
34	217	9411.0	05	1644	SLA	U2
35	319	9424.2	08	0805	SLA	U4
35	436	9423.6	10	0923	SLA	U4
35	465	9422.6	10	1320	SLA	U4
36	433	9419.0	10	0848	SLA	W4
37	057	9412.6	02	1003	SLA	X2
37	098	9413.0	02	1647	SLA	X2

Category	Collect #	RF	Date	Time	Location	ID
37	150	9413.6	04	0937	SLA	X2
37	222	9414.3	06	0829	SLA	X2
38	393	9403.4	09	0855	SLA	X4
38	394	9403.4	09	0855	SLA	X4
39	163	9416.1	04	1151	SLA	Z2
39	170	9416.2	04	1408	SLA	Z2
39	182	9417.7	05	0811	SLA	Z2
39	183	9417.7	05	0811	SLA	Z2
39	610	9413.2	12	1440	SLA	Z2
39	619	9412.8	12	1544	SLA	Z2
40	564	9408.3	11	1544	SLA	Z4