ERDC WQTN-AM-05
March 2000

Spatial Interpolation Techniques for Water Quality Analysis

by Perry LaPotin[1] and Robert Kennedy[2]

PURPOSE: With the development of high-frequency data acquisition systems, water quality (WQ) measurements may be collected in near real-time along preprogrammed transects. The data are commonly linked to GPS coordinates and analyzed to determine spatial dependencies. This technical note examines techniques for analyzing these data with spatial interpolation methods. The interpolation is used to portray the data as contours that may be viewed for each respective WQ parameter by surface position and column profile. This technical note examines the statistical methods required to perform optimal interpolation using discriminant theory. The approach is quite general, and is shown to yield optimal separation of points within class contours. Optimality is described from a statistical basis using error probability measures.

BACKGROUND: Assessment of limnological data is closely related to the adequacy, precision, and fidelity of the statistical sample. However, in most applications, a complete high-density stratified sample is not feasible due to logistical requirements and cost considerations. For this reason, a limited (focused) sample is acquired within select strata as a means to estimate the overall spatial variability. More often than not, questions arise concerning the applicability of these data, i.e. how representative are these measures gathered at sparse postings, and how well do they explain the overall features and patterns in a complex lake system? Furthermore, how can these measurements be organized for proper application within numerical models?

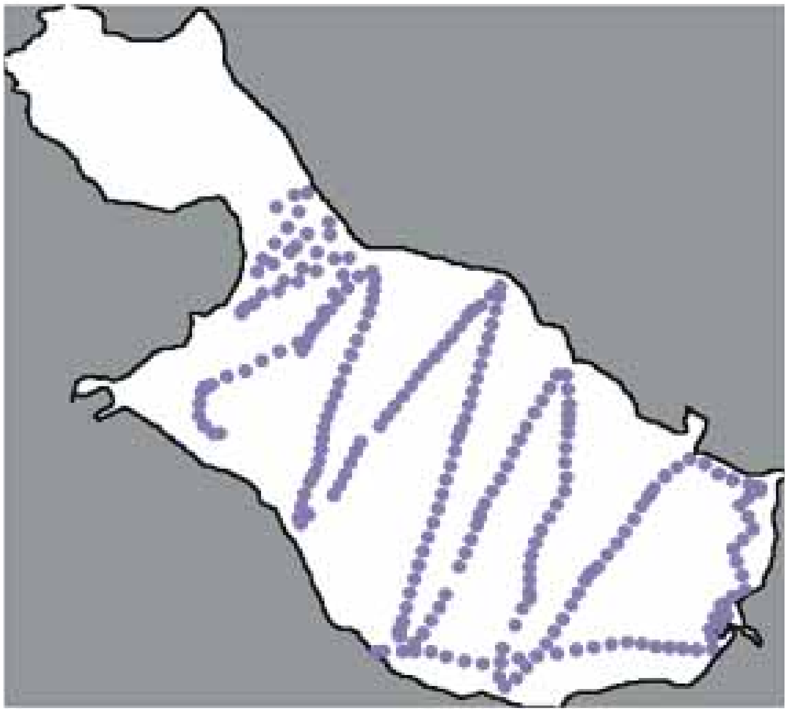
In recent applications (Kennedy, Meyer, and Cremeans 1999), high-frequency geo-referenced data have been acquired at 1- to 2-sec intervals along selected transects. The transects approximately span the full lake system in a crosswise pattern. Figure 1 is an example of this type of data application (GPS tracking) for San Jose Lagoon, San Juan, Puerto Rico.[3] In this application, an onboard sensor package and surface pump were used to acquire WQ data along the displayed path. Using general techniques (Cressie 1998), an approximate contour may be generated that displays the spatial variability of the sample for selected parameters. An example of this contouring methodology is shown in Figure 1b using primary data acquired along the GPS path portrayed in Figure 1a. As provided in Figure 1b, the fluorescence contour is only one possible representation of the data. No summary measures are provided for the accuracy and precision of this representation. Indeed, most contouring methodologies provide few measures for determining the proper contouring of the posted data (Yuhas, Goetz, and Boardman 1998).

In Figures 2a-2h, a second example data set is provided for data extracted from the GPS transect shown in Figure 2a. These data were acquired for chlorophyll at J. Strom Thurmond Lake, a U.S.

1 United Imaging Corporation, Lyme, NH.
2 Environmental Laboratory Engineer Research and Development Center, Vicksburg, MS.

3 The San Jose Lagoon data include paired observations of GPS position and fluorescence, a measure of chlorophyll, collected at the surface from a boat repeatedly traversing the lake over its length.

a. GPS tracking for San Jose Lagoon, San Juan, Puerto Rico. Information was acquired using an onboard sensor system with co-registered GPS data to mark each acquisition



b. WQ contours derived from the GPS track shown in Figure 1a. This plot shows one potential contour representation for the acquired data. Additional contours may be derived using alternative methodologies such as nonlinear weighting, and gradient contouring.

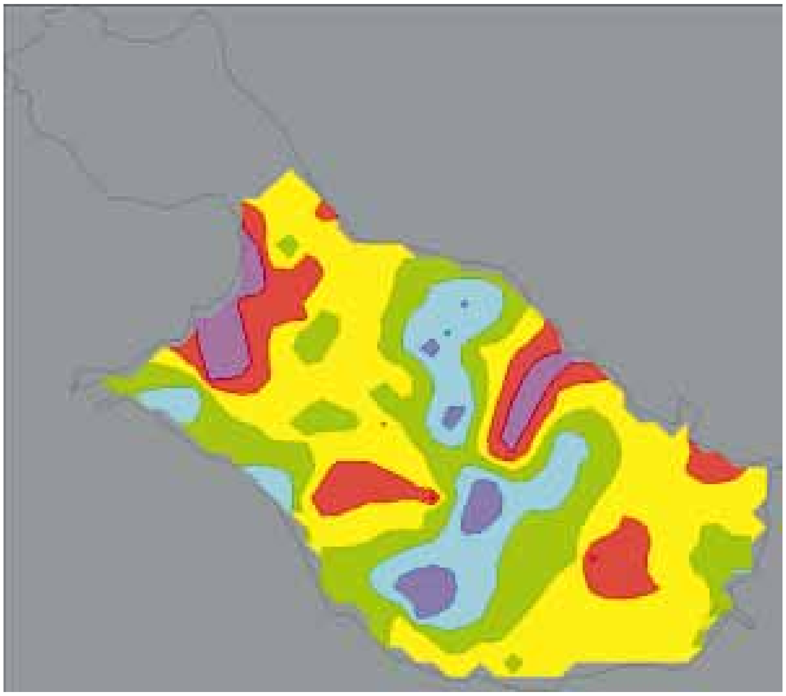
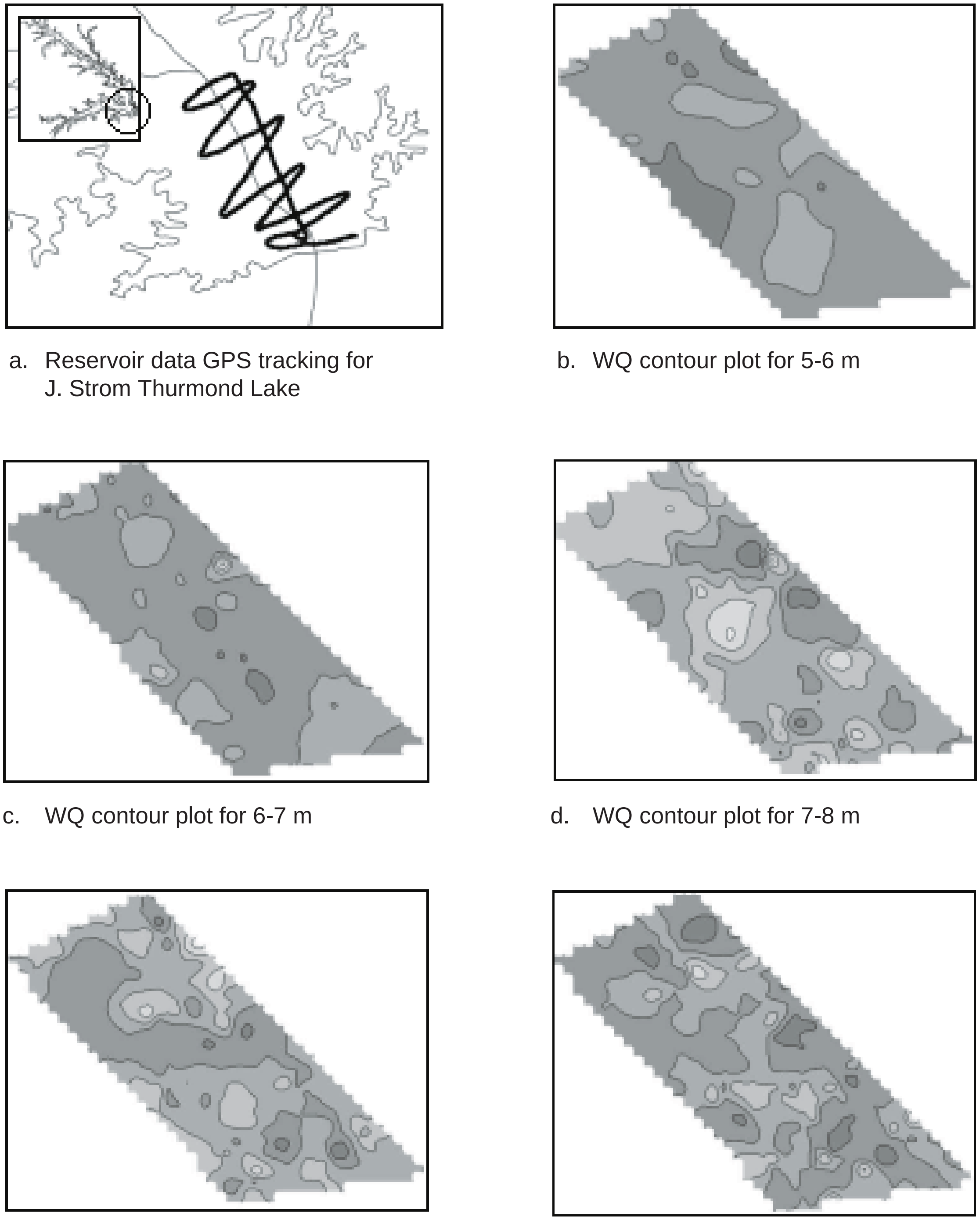


Figure 1. Example data application and contouring methodology for San Jose Lagoon, San Juan, Puerto Rico



a. Reservoir data GPS tracking for J. Strom Thurmond Lake

b. WQ contour plot for 5-6 m

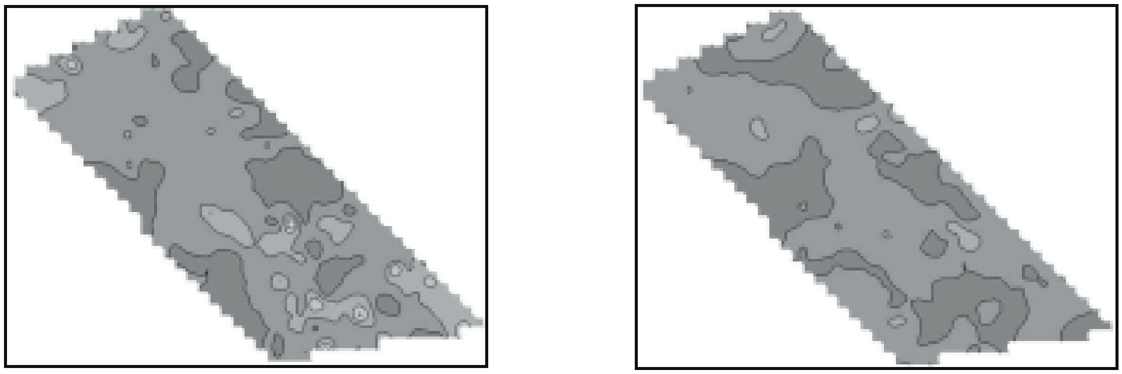
c. WQ contour plot for 6-7 m

d. WQ contour plot for 7-8 m

e. WQ contour plot for 8-9 m

f. WQ contour plot for 9-10 m

Figure 2. Example data set for chlorophyll at J. Strom Thurmond Lake (Continued)



g. WQ contour plot for 10-11 m

h. WQ contour plot for 11-12 m

Figure 2. (Concluded)

Army Corps of Engineers hydropower reservoir located on the Savannah River near Augusta, GA. Within this diagram, the spatial distribution for chlorophyll in μg/l is shown at varied lake depths using data acquired by a towed, in situ sensor package (Hains and Kennedy 2000). The respective contours by depth are shown in Figures 2b-2h. The plots indicate that the number of contours varies considerably by depth profile. The 5- to 6-m acquisition contains few iso-contours with respect to equivalent data acquisition at the 9-10 m level. Similarly, contours vary across nearly equivalent depth profiles. A visual contrast of the 6- to 7-m acquisition versus the 7- to 8-m acquisition displays significant deviations in the number of contours and their respective position on the contour plot. The following questions remain:

a. During the acquisition phase, how can the lake system be optimally stratified for the detection of WQ parameters (chlorophyll, Secchi depth, temperature, dissolved oxygen, pH, etc.)? In other words, where (within the confines of the lake) should the towed vehicle go so that sensors (acquiring information at multiple depth profiles) are optimally used to acquire data that best describe the overall variability in WQ phenomena.

b. Given an optimal sampling strategy in item *a.* above, what statistical approach should be used to group or contour the samples into optimal classes? In addition, how should these spatial patterns be displayed (and what related statistical measures should be provided) to show the accuracy and precision of the spatial representation?

In Figure 1b and related contours shown in Figures 2b-2h, homogeneous classes are used to portray the data. This implies that a single color (class) is always assigned to a contour interval. For example, in the chlorophyll plots of Figures 2b-2h, the analysis indicates a nearly homogeneous distribution of chlorophyll between each contour line. The full range of values varies between 0 μg/l and 12 μg/l as shown in Figure 2a. The "lumping" of a continuous variable (chlorophyll) into discrete classes (contours) is the subject of this technical note. Specifically, what discriminate functions can be applied to perform this separation so that the portrayal of the data is optimal? This question is often difficult to answer since it is tightly coupled to the sampling design as a related consideration posed in item *a*. above. Indeed, the separation function (discriminate function) is the subject of ongoing research in image processing and numerical analysis (Clark et al. 1998). This technical note examines techniques for the separation of WQ data into iso-contours. The general methodology is discriminate analysis, since this approach is generally available in numerical analysis, and converges on an optimal solution for *a priori*, well-defined constraints. However, our approach is to consider various deviations from the classical discriminate methodology so that we may define a more optimal approach for WQ analysis.

METHODOLOGY: Traditional contour analysis uses discrete data at prescribed x,y,z locations to generate a continuous profile. The methodology is quite general, and often varies to a considerable degree by software application (Kruskal 1996). The standard approach uses simple interpolation to derive a class structure. In this method, the z azimuth is neglected, and the interpolated distance between two WQ sample points $P_1(x_1,y_1)$ and $P_2(x_2, y_2)$ is:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

From standard Euclidean geometry, the slope of the connection is simply the derivative approximation $\hat{m}$:

$$\hat{m} = \frac{y_2 - y_1}{x_2 - x_1} = \tan\theta \tag{2}$$

The skeleton geometry is shown in Figure 3 for two sample points $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$.

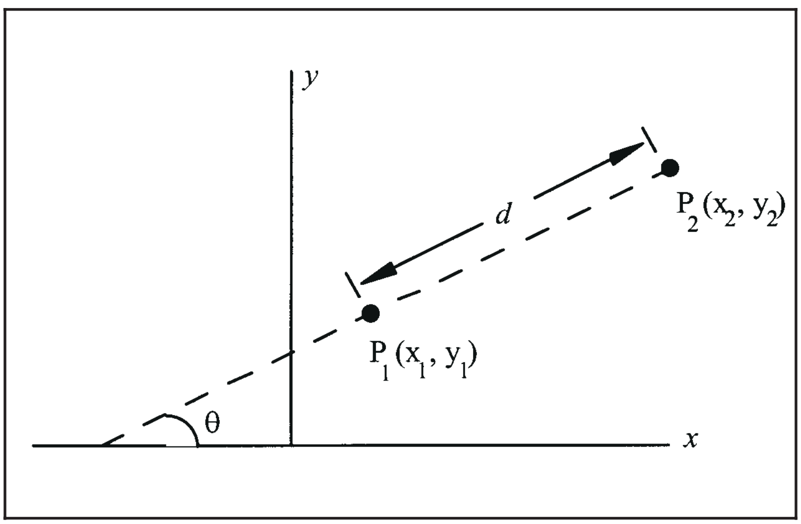


Figure 3. Simple interpolation between two points P_1 and P_2 in two dimensions

For planar data (two dimensions) that include a third measure (either the z azimuth for PS data or a single WQ parameter), the geometry shown in Figure 3 may be extended to the form shown in Figure 4. Once again, the diagram uses simple Euclidean geometry for three-dimensional data. The respective equations for distance and slope are provided in Equations 3a-3c and Equations 4a and 4b, respectively.

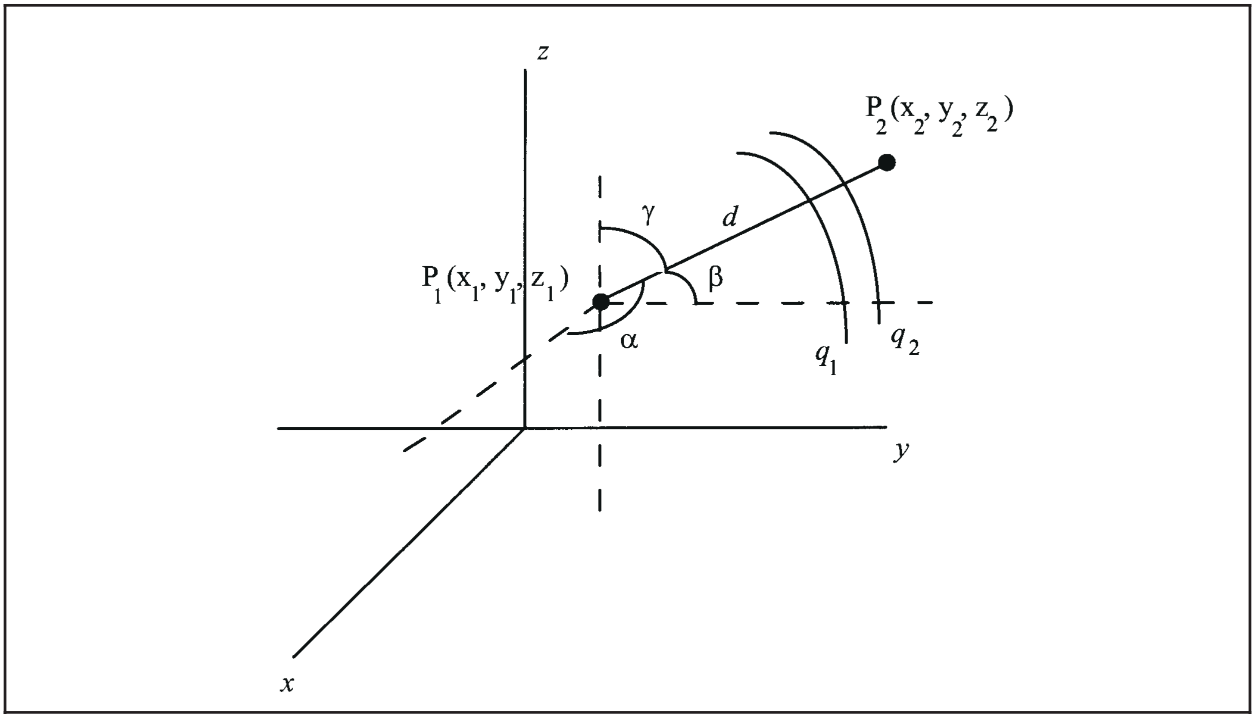


Figure 4. Simple interpolation between two points P_1 and P_2 in three dimensions. Within the contouring algorithm, a separation line *q* is created between the two points. The line creates a class boundary for the variable z. In standard contouring applications, the variable z represents an independent parameter that is functionally dependent upon x and y. For the case of WQ analysis, the variable z may represent chlorophyll or dissolved oxygen or indeed any independent measure. In this illustration two perspective separation lines are provided, q_1 and q_2. The lines are actually curvilinear surfaces depending on the number of points in the sample. As distinct values of z are added, the curvature increases to form a polygonal mapping that separates the data points.

$$\hat{m} = \frac{y_2 - y_1}{d} = \cos\beta \quad (3a)$$

$$\hat{l} = \frac{x_2 - x_1}{d} = \cos\alpha \quad (3b)$$

$$\hat{n} = \frac{z_2 - z_1}{d} = \frac{\eta_2 - \eta_1}{d} = \cos\gamma \quad (3c)$$

Note that the following constraint remains in effect for all of *R*, where *R* indicates the plane of real numbers.

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1 \quad (4a)$$

or

$$\hat{l}^2 + \hat{m}^2 + \hat{n}^2 = 1 \quad (4b)$$

In Equations 3a-3c, direction cosines are provided to show how a standard interpolation algorithm finds a new path through the data. Each new path is a prospective class boundary defined by a separation line (a polygon perimeter) q_1 or q_2. In Equation 3c, the variable η denotes a representative water quality measurement, so each algorithm must locate an intersection point on the distance line d to create a separation line. Some comments on this process are in order:

a. The creation of a distance metric between measurements along the (x,y) plane is strictly geometric and no stochastic properties are required.

b. The position of the separation line may be derived based upon some generic-weighting scheme. For many algorithms, an average surface is created that separates the water quality measurement at P_1 from P_2 using an equal-distance function. Other algorithms use an inverse weighting function (similar to laws that govern gravitational attraction of two bodies of unequal density). In all cases, an a priori method is selected to create the position and mapping of the separation line. The separation line is connected to form a closed (non-degenerate) polygon. These polygons are the iso-contours shown in Figure 1b and Figures 2b-2h.

Some problems with this approach may be quickly identified:

a. Since the technique is geometrically derived, probability measures cannot be directly calculated from the contour procedure. Restated, the method finds a geometric path through the data that is likely to be nonoptimal with respect to stochastic properties.

b. Interpolation does not adequately manage clustered data or sparse data samples. The approach is quite good for equal area samples (e.g. a uniform grid of sample data evenly distributed across the lake surface), but is poorly adapted to oblique transect data along a complex path (e.g., Figures 1a and 2a - GPS trackings).

c. The technique does not yield stochastic measures of accuracy or precision with respect to the final map product. Restated, how accurate is the iso-contour representation of the discrete data and does the iso-contour plot truly display the spatial variability in the underlying WQ parameter (chlorophyll, Secchi depth, turbidity, etc.)?

As an alternative methodology, discriminate analysis may be considered as a stochastic technique for the creation of an *optimal* separation line. From the discriminate function, classification and missclassification errors may be derived to determine the accuracy of the iso-contour plot.

Discrimination. Discriminate analysis and classification are multivariate techniques concerned with separating distinct sets of objects (measurements) and with allocating new objects (observations) to previously defined groups. As a separation procedure, it is often employed on a one-time basis to investigate the differences when causal relationships (or spatial dependencies) are not well understood. Hence, the immediate goals are to:

a. Describe graphically (in three or fewer dimensions) or algebraically (in higher dimensions), the differential features of objects (measurements) from several known collections (populations). This approach attempts to derive discriminants whose numerical values are such that the collections are separated as much as possible.

b. Sort observations into two or more labeled classes (contours). The emphasis is on deriving a rule that can be used to optimally assign a new object to the labeled class. It is hoped that the rule is helpful in determining a causal mapping that assists the researcher in the determination of how and why this separation has occurred. This may or may not be possible, or practical, depending upon the data set and the level of interpretation.

The general two-group discrimination problem may be characterized as follows: on the basis of measurements alone, a WQ sample must be assigned to one of two groups. It is assumed that the two groups are the only possible choices and that an assignment must be made. This general model conforms to the k_2 class discriminate problem as used within standard image processing and conformal GIS modeling of multispectral data (Mazer et al. 1997). After the assignment procedure has been established, it may also be desired to estimate the proportion of pixels (clusters) in each group (class) that will be “wrongly” classified (missclassification error). One approach to the classification problem is to assign a linear combination of the observations to form a single number. Then an individual observation is classified into one group (or the other group) depending on this number (the threshold). In this capacity, a single digit is used as a decision measure for assigning any observed WQ measurement into a specific contour. Several means of estimating error rates for a given general discriminate function will be considered. The derivation is as follows:

Let $x_k = (x_{1k}, x_{2k}, ..., x_{pk})'$ denote the kth sample value of a
p-dimensional multivariate column vector k=1,2, .., $n_1 + n_2 = m$.

In this notation, the vector x holds p-4 WQ variables. The remaining four variables are likely to be spatial or temporal values. For example, the observations in positions x_{1k}, x_{2k}, and x_{3k} may be the respective (x,y,z) GPS triplets that mark the sample location. Similarly, a single position is set for the time stamp of the observed WQ sample (e.g. in position x_{4k}). Since the orientation is quite general, the specific location of these positional/temporal covariates is not important and will not affect the derivation of the optimal separation line.

The within-group sample covariance between the *i*th and *j*th variables will be denoted by sij. Mean values from the group samples are denoted by $\underline{x}_1 = (x_{11}, x_{21}, ..., x_{p1})'$ and $\underline{x}_2 = (x_{12}, x_{22}, ..., x_{p2})'$.

The general method for classification is to find a linear combination of pixels that would maximize the difference between groups *relative* to the standard deviation *within* groups. This leads to the choice discriminate function: $S^{-1} (\underline{x}_1 - \underline{x}_2)$ where *S* is the sample covariance matrix. Then:

Classify as group 1 if $x'S^{-1} (\underline{x}_1 - \underline{x}_2)$ is greater than some constant C,

and (5)

Classify as group 2 otherwise.

If the density functions are denoted by f_1 and f_2, the likelihood ratio criterion leads to the rule: assign the observation to group 1 if $f_1 / f_2 > k$ and otherwise to group 2. The choice of k depends on the relative costs of misclassification, and the proportion of each group in the general population.

If f_1 and f_2 are multivariate normal distributions with the same covariance matrix Σ, but different means μ_1 and μ_2, it may be shown that the likelihood ratio criterion is equivalent to classifying as group 1 given the discriminate function:

$$D_t(x) = x' \Sigma^{-1}(\mu_1 - \mu_2) - 0.5(\mu_1 + \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2) \quad (6)$$

The first part of this expression is the theoretical analogue of the function suggested by standard multivariate logic. If the population parameters of the distribution were known, this expression would be optimal. However, the population parameters are generally unknown, so it is necessary to estimate them from within a sample volume for selected data. Thus, the discriminate function to be concerned with is:

$$D_s(x) = x' S^{-1}(x_1 - x_2) - 0.5(x_1 + x_2)' S^{-1}(x_1 - x_2) \quad (7)$$

To simplify the notation in Equation 7, let $y = (x_1 + x_2)$ and $z = (x_1 - x_2)$. Then:

$$D_s(x) = x' S^{-1} z - 0.5 y' S^{-1} z \quad (8)$$

The probabilities of misclassification will be estimated based on the sample discriminate function. If the parameters of the distribution were known, the problem would be trivial, and reduce to $P_1 = P_2 = \phi(-\delta/2)$ where $\delta_2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$ is Mahalanobis distance, and ϕ is the cumulative normal distribution. This represents a limiting value, which cannot be improved upon.

Error Estimation. Error estimation techniques may be divided into two classes: those using a sample to evaluate a given discriminate function, and those using properties of the normal (Gaussian) distribution. The former may be considered empirical methods, while the latter are dependent on the normality of the distribution for their underlying validity.

The H Method. If the initial samples are sufficiently large, a subset of observations can be chosen from each group, a discriminate function can be computed from them, and the remaining information can be used to estimate the error probabilities. The number of errors in each group will be binomially distributed with probabilities P_1 and P_2. After these estimates have been obtained, the discriminate function can be recompiled using the entire sample. There are several drawbacks to this method:

a. In many WQ applications, large samples are not available (too costly to acquire). This is particularly true in complex ecosystems where a limited number of ground truth classes are known, and each in situ sample is expensive to acquire with precise GPS positioning.

b. The discriminate function that is evaluated is not the one that is used. There may be a considerable difference in the performance of the two.

c. There are problems connected with the size of the holdout sample. If it is large, a good estimate of the performance of the discriminate function will be obtained but the function is likely to be a poor classification measure. If the holdout sample is small, the discriminate function will be better, but the estimate of its performance will be highly variable.

d. This method is quite uneconomical with data. A larger sample than is necessary to obtain a good discriminate function must be selected to estimate performance. For these reasons, the holdout method should not be used for applications that require costly data acquisition. This empirical method is referred to as the *H* method or holdout method.

The R Method. The second commonly used technique for discrimination is referred to as the *resubstitution* method or *R* method. In this process, the sample used to compute the discriminate function is also used to estimate the error. Approximate values of such error may be obtained in two ways. First, taking the distribution of x to be f_a xdx in population *a*, the distribution of L_x (the discriminate function) may be extracted from it and from that the probability that $L_x < L_o$. Unfortunately, the integrals needed to evaluate this probability measure are frequently too difficult for empirical calculations; and, in addition, if we have extracted the wrong form of the distribution f_a, the value of the error E_a may be quite wrong. Alternatively, if a sample is taken from the population *a*, and frequency of $L_x < L_o$ in this sample is determined, then the estimate for f_a can be improved. For this application we have a sample, namely, the one that was used in working out the form of the distribution f_a x. However, it is necessary to keep in mind that the sample used to get E_a will not have a binomial value: sqrt{ $E_a(1-E_a)/n_a$ }, although it will probably not be greatly different in form or magnitude.

We have found that the *H* and the *R* methods can quickly lead to misleading error estimation probabilities. If the ground-truth sample used to compute the discriminate function is not large, each method gives too optimistic an estimate of the probabilities of misclassification. The main point, it seems, is not that the estimate does not have the binomial standard error, but that it can be a badly biased measure depending upon sample magnitude. Within the *R* method, the probability of misclassification P_1 may be written as:

$$P_1 = P\left\{ t < \left(-\mu_1 + 0.5y\right)' S^{-1} z / \mathrm{sqrt}\left(z' S^{-1} \Sigma \ S^{-1} z\right) \right\} \quad (9)$$

where *t* is an index for the central t-distribution.

If μ_1 and Σ are replaced by $\underline{x}_1$ and S, for normally distributed variables, the estimate of P_1 (or P_2) is $\phi(-\Delta/2)$ where $\Delta^2 = z' S^{-1} z$ is a secondary Mahalanobis distance measure.

If the degrees of freedom are large (large sample size), this is a fairly accurate estimate of P_1, since Δ^2 is consistent for δ^2. If the degrees of freedom are not large (small sample poorly distributed within the lake system), this may be badly biased and give too favorable an impression of the probability of error.

EXAMPLE APPLICATION: In Figure 5, sample GPS postings are shown for WQ data acquired at West Point Lake, GA. A complete discussion of this data set and corresponding image data are provided in LaPotin and Kennedy (2000) and Kennedy et al. (1994). As shown, information was acquired along a principal axis (from lower left to upper right). This portrayal is partially due to the scaling selected for this display (i.e., y scale from 118.4 to 119.8 latitudinal units and x scale from 305.1 to 305.9 longitudinal units), and also reflects the specific subgroup of data (the portion of the complete GPS survey under analysis). In Figures 6a-6d, contour plots are provided for Secchi depth acquired at the GPS positions shown in Figure 5. Therefore, each contour plot in Figures 6a-6d has an identical x,y axis and spatial scaling identical to that shown in Figure 5.

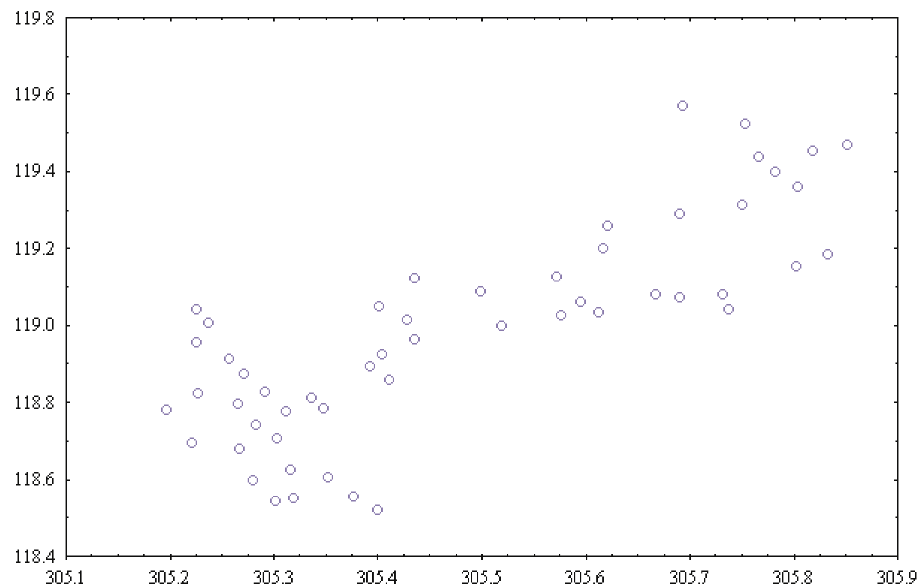
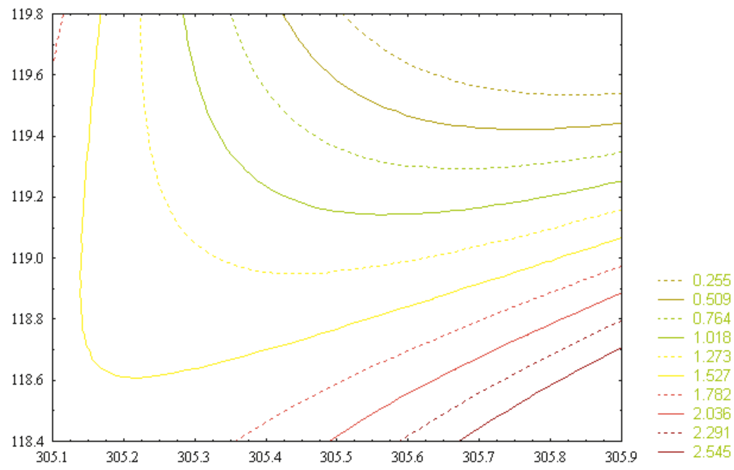


Figure 5. GPS postings for the West Point Lake, Georgia, WQ assessment. In this figure, data are judiciously selected to distribute information along a single primary axis. The data set is intentionally sparse in the upper left and lower right corners to test the respective contour methods (in the absence of available data)

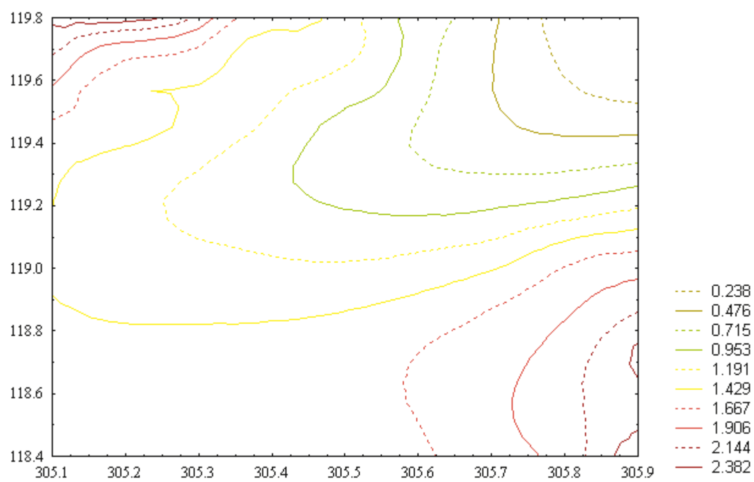
In this trial, data were selected along a principal axis to observe the behavior of four analytical methods:

- Quadratic interpolation.
- Discriminate interpolation using a two-standard deviation threshold.
- Discriminate interpolation using the *H*-method to estimate the threshold level.
- Discriminate interpolation using the *R*-method to estimate the threshold level.

The primary variable of interest is Secchi depth acquired by Kennedy et al. (1994) at 52 sample locations on June 8, 1991. In Figure 6, the spatial distribution of Secchi depth is provided using quadratic interpolation to build respective contours. This method is the general “default technique” used in most commercial software applications (Richards 1999; Smith, Pyden, and Cole 1999). The

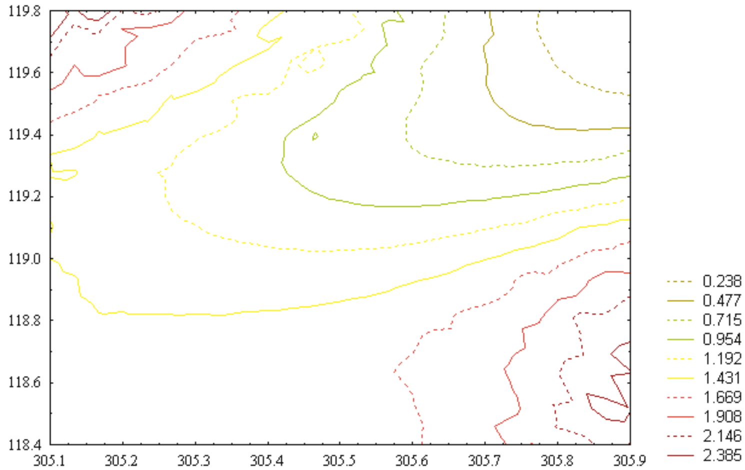


a. Quadratic interpolation. Contours are shown for Secchi depth at the GPS postings provided in Figure 5. This analytical method uses a linear segmentation similar to that shown in Equation 4 with an additional higher order quadratic fit. The result is usually hyperbolic or parabolic depending on the data set and spatial distribution

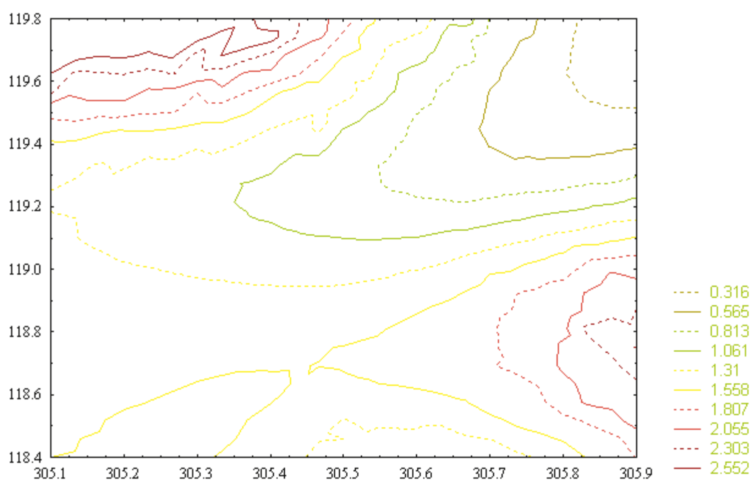


b. Discriminate interpolation. Contours are shown for Secchi depth at the GPS postings provided in Figure 5. This analytical method uses a single threshold level to separate the contours into groups (two at a time). The grouping logic shown in Equation 5 is repeatedly applied until all data are placed within a contour

Figure 6. Interpolation methods (Continued)



c. Discriminate interpolation - *H* method. Contours are shown for Secchi depth at the GPS postings provided in Figure 5. This analytical method uses a subset of observations to calculate the error probabilities for each class. The technique is faster than the *R* method, since a fraction of the observations are required to determine the discriminate threshold



d. Discriminate interpolation - *R* method. Contours are shown for Secchi depth at the GPS postings provided in Figure 5. This analytical method uses the full set of observations to calculate the error probabilities for each class. The technique is slower than the *H* method, since all observations are required to determine the discriminate threshold

Figure 6. (Concluded)

method is a simple extension of the first-order linear separation line shown in Figure 4. The algorithm is known to be highly sensitive to sparse data and extreme outliers. In this example, contours are generated as elliptical surfaces. The Secchi depth is a maximum along the two outer contour lines and is minimized toward the center (primary axis) of the data.

The sharp contours are generated by second-order quadratic fit of the data. The result is a bending profile that is (more or less) replicated by each iso-contour. This bending to the extreme of the plot usually indicates that a higher-order fit is justified for the data and that a simple linear segmentation would underestimate most patterns in the secchi depth.

In Figure 6b, a primary discriminate plot is provided that separates Secchi depth into pairwise classes. The data are grouped into a contour if the weighted mean difference $x'S^{-1} (\underline{x}_1 - \underline{x}_2)$ exceeds a threshold C. The threshold is a 2-sigma deviation to indicate strong separation at the 10-percent level or above. The contours produced by this method are more regular (and well-defined) when compared to the quadratic interpolation. Since the threshold varies with the standard deviation, and the separability is optimal with respect to the difference in mean Secchi depth, this technique will (for nearly all data sets) out-perform the standard quadratic interpolation technique.

In Figure 6c, a discriminate plot is shown for secchi depth using a subset selection of observations to calculate the discriminate threshold and related error probabilities. The technique produces contours that parallel the standard discriminate method in Figure 6d with additional detail for regions. Additional separations are shown where the data are sparse. This is due to the fact that these regions (upper left and lower right) are estimated from data along the main axis using subset sampling. This method judiciously uses limited data, but may over- (under-) estimate the true contours since all data are not used to create the discriminate function and related error probabilities.

Figure 6d is a discriminate plot for Secchi depth using the resubstitution method. In this technique, the sample used to compute the discriminate function is also used to estimate the error. For this reason, the error probability is unbiased and closely coupled to the discriminate threshold level.

CONCLUSIONS: Recent developments in electronic WQ monitoring instrumentation and the availability of GPS technology have, and will continue to, dramatically increase the capabilities of limnological researchers and water quality managers to acquire detailed information on environmental variability in both space and time. Kennedy, Meyer, and Cremeans (1999) and Hains and Kennedy (2000) recently demonstrated such capabilities in rivers and reservoirs, respectively. In each case, the number of sampling points was greatly increased over the number possible using traditional sample collection methods. However, the ability to discern pattern and trends, or to compare patterns across both space and time will require that the interpolation method accounts for the inherent variability associated with these data.

In this technical note, spatial interpolation methods are compared using geometric models to calculate the separation surface. For two points, the separation q is a single line that uniquely defines two classes for the posted data. For three or more points, direction cosines are calculated to describe a higher order separation surface. Using geometric separation methods (linear, quadratic, and polynomial contouring), a separation surface is calculated based upon an "equal-area" criterion. This approach is intuitive, but does not utilize the stochastic relationships within the data. Furthermore, these techniques implicitly assume a gridded equal area distribution of the sample points. Strong sample clustering and sparse sample regions significantly bias the resultant contours. As an alternative methodology, a discriminate function may be derived that utilizes the weighted mean difference between observed groups (iso-contours). Using these techniques, an optimal discriminate surface may be created that minimizes the distributed error (minimum Mahalanobis Distance). Three related strategies are proposed: (a) single technique for utilizing the standard discriminate model using a fixed 2σ threshold; (b) a single technique that optimizes the processing speed by repeated subset selection of the full data set; and (c) a method for reapplication of the data to calculate the discriminate function and the related error probabilities. For the West Point Lake trials, the three discriminate methods produced consistent representations for the spatial delineation of Secchi depth. Each contour differed in the manner (weighting) applied to missing data; however, all surfaces were similar in their representation. Unlike the quadratic representation, few contours resulted in purely parabolic or hypergeometric surfaces. This suggests model closure and proper application of the weighted discriminate surface.

POINTS OF CONTACT: For additional information, contact Dr. Robert H. Kennedy, Engineer Research and Development Center, Waterways Experiment Station, Vicksburg, MS, (601) 634-3659, *kennedr@wes.army.mil* or the managers of the Water Quality Research Program, Dr. John Barko, (601) 634-3654, *barkoj@wes.army.mil* and Mr. Robert C. Gunkel, Jr., (601) 624-3722, *gunkelr@wes.army.mil*. This technical note should be cited as follows:

> LaPotin, P., and Kennedy, R. (2000). "Spatial interpolation techniques for water quality analysis," *Water Quality Technical Notes Collection* (ERDC WQTN-AM-05), U.S. Army Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/elpubs/wqtncont.html*

REFERENCES

Clark, R. N., Swayze, G. A., Gallagher, A. J., and Kruse, F. A. (1998). "Mapping with imaging spectrometer data using the complete band shape least-squares algorithm simultaneously fit to multiple spectral features," *Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, JPL Publication 91-28.

Cressie, N. (1998). *Statistics for spatial data*, 2nd ed., Wiley, New York.

Hains, J. J., and Kennedy, R. H. (2000). "Investigating water quality with a towed sensor array," *Water Quality Technical Notes Collection* (ERDC WQTN-AM-06), U.S. Army Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/elpubs/wqtncont.html*

Kennedy, R. H., and Walker, W. W., Jr. (1990). "Reservoir nutrient dynamics." *Perspectives in reservoir limnology*, K. W. Thornton, B. L. Kimmel, and F. E. Payne, ed., John Wiley and Sons, New York.

Kennedy, R. H., Hains, J. J., Ashby, S. L., Jabour, W., Naugle, B., and Speziale B. (1994). "Limnological assessment of West Point Lake, Georgia," Technical Report EL-94-6, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Kennedy, R. H., Meyer, R. M., and. Cremeans, W. L. (1999). "Methods for collecting high-frequency, geographically referenced in-situ water quality information." *Water Quality Technical Notes Collection* (WQTN AM-04). Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/elpubs/wqtncont.html*

Kruskal, C. (1996). "Searching, merging and sorting in parallel computation," *IEEE Transactions on Computers* C32(10).

LaPotin, P., and Kennedy, R. H. (2000). "Development of unsupervised classification techniques for water quality analysis," *Water Quality Technical Notes Collection.* (ERDC WQTN-AM-07), U.S. Engineer Research and Development Center, Vicksburg, MS. *www.wes.army.mil/el/elpubs/wqtncont.html*

Mazer, A. S., Martin, M., Lee, M., and Solomon, J. E. (1997). "Image processing software for imaging spectrometry data analysis," *Remote Sensing of the Environment* 24(1).

Richards, J. A. (1999). *Remote sensing digital image analysis. An introduction*, 3rd ed., Springer-Verlag, Berlin.

Smith, C., Pyden, N., and Cole, P. (1999). *ERDAS field guide.* Edition 9. ERDAS, Inc., Atlanta, GA.

Yuhas, R. H., Goetz, A. F. H., and Boardman, J. W. (1998). "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," *Summaries of the Third Airborne Geoscience Workshop*, JPL Publication 92-14, v1.