PEER-REVIEWED ARTICLE

Parametric and Nearest-Neighbor Methods for Hybrid Classification: A Comparison of Pixel Assignment Accuracy

Perry J. Hardin

Abstract

Nearest-neighbor classifiers have not been widely used for pixel assignment, probably because their computational requirements make them too slow for practical application to large images. Regardless, when properly specified, the nearest-neighbor classifier is a Bayesian classifier, and does not require conditions of multivariate normality as a prerequisite for optimum pixel assignment. In this study, six nearest-neighbor classifiers and four parametric classifiers are applied to six Landsat images embracing a broad variety of land cover. The accuracy of the parametric- and neighbor-based methods is compared for significant differences in the assignment of pixels to spectral classes. In a majority of hybrid experiments, classifiers predicated on spectral neighborhoods were significantly superior to parametric classifiers for pixel assignment when training sample proportions matched the true population proportions. In experiments where this condition was violated, there was no clear advantage to choosing a neighbor-based classifier in preference to a linear discriminant function employing prior probabilities.

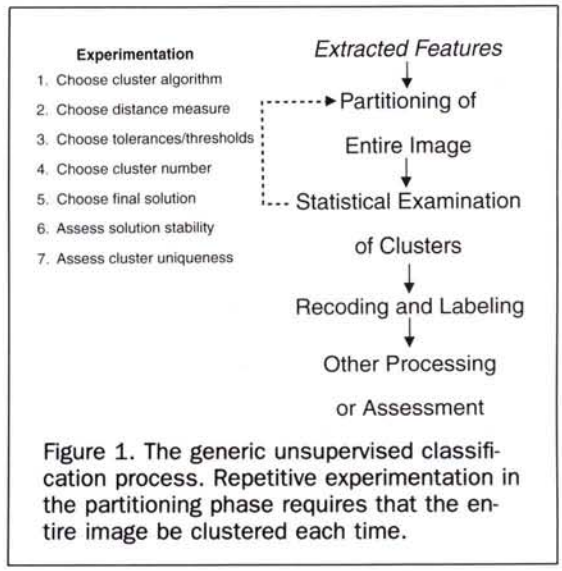

Figure 1. The generic unsupervised classification process. Repetitive experimentation in the partitioning phase requires that the entire image be clustered each time.

Introduction

In simple form, unsupervised spectral classification can be condensed into four steps. As discussed by Campbell (1987), these include (1) predetermining the range of spectral classes desired from the clustering process, (2) selecting a clustering algorithm, (3) assigning image pixels to the clusters through application of the selected algorithm (i.e., partitioning), and (4) placing the resulting clusters into an information class. After the first two steps are complete, there is little interaction between the analyst and the clustering algorithm, and the results are an "objective" classification (Campbell, 1987).

In applications where the utility of a classification is more important than its objectivity, simple unsupervised classification may be conducted in three broad phases. As shown in Figure 1, these phases include (1) partitioning the entire image using a selected clustering algorithm, (2) statistically examining the clustering results, and (3) recoding the final spectral classes into some project-driven set of information classes. In practice, the partitioning and statistical examination phases might be performed several times in order to find an acceptable solution.

However, because the correspondence between the spectral classes and desired information classes is frequently low, the practical value of a totally unsupervised clustering may be inadequate for the problem under consideration. Furthermore, the computer time required to cluster large multiband images makes repetition of the clustering phase time consuming, precluding a flexible exploratory approach involving significant analyst interaction.

Because true unsupervised classification can produce unusable results, can be time consuming, and can limit analyst interaction, many researchers select a semisupervised or hybrid approach in preference to the completely unsupervised method outlined above (Campbell, 1987). In one variant of hybrid classification (see Figure 2), a random or systematic sample is extracted from the image and submitted to the clustering algorithm. Descriptive statistics are calculated for each resulting cluster, and a classifier such as maximum likelihood is then used to assign all the pixels in the image to a spectral class.

Speed of analysis, flexibility, and more potential for analyst-process interplay are the primary advantages of the hybrid approach over its general unsupervised counterpart.

Department of Geography, 690 SWKT, Brigham Young University, Provo, UT 84602.

Photogrammetric Engineering & Remote Sensing, Vol. 60, No. 12, December 1994, pp. 1439–1448.

0099-1112/94/6012-1439$3.00/0

Because a sample is being clustered instead of the entire image, multiple iterations through the partitioning-examination phases can be repeated more rapidly. In addition to the experimentation performed at the partitioning stage, experimentation is also possible during the classifier training phase. For example, clusters can be aggregated or split manually and tentatively labeled. Several classifiers can be tested and adjusted to determine which achieves the highest possible accuracy. Should this experimentation reveal a poor cluster solution, the partitioning phase can be reconducted without reclustering the entire image. Finally, even when combined with the execution time required to extract and cluster a sample, the time required to classify an entire image with a parametric classifier in the hybrid process is usually less than the time required to cluster the image in its entirety.[1] Because of these advantages, nearly all commercial software gives an analyst the option to perform unsupervised classification using some hybrid method.

The comparison of hybrid classification with supervised classification is inevitable because both utilize a pixel assignment device at some stage of the classification process. However, the pixel assignment step of supervised classification is maximized to insure the highest accuracy in the assignment of image pixels to *a priori* informational classes, whereas the goal of the pixel assignment step in hybrid classification is to assign all the image pixels to spectral classes, leaving the issue of final informational classes unresolved until later in the project.

The most widely accepted pixel assignment device for supervised and hybrid land-cover mapping projects is the maximum-likelihood classifier. Parametric maximum-likelihood classifiers are a family of methods which statistically summarize the training data (or cluster classes) and assign pixel labels through the application of probability theory. In contrast to the popular maximum-likelihood classifier, nonparametric classifiers predicated on spectral neighborhoods have found little application in remote sensing. Unlike their parametric counterparts, spectral neighborhood classifiers do not summarize the training (or cluster) classes prior to the pixel assignment step, but instead store all the training (or cluster) pixel brightness values (BVs) in memory. In the pixel assignment step itself, an unlabeled pixel is given a class label by "taking a vote" among its nearest neighboring training (or cluster) pixels in feature space.

[1]In the author's experience, most clustering algorithms employed in image classification execute in time approximately proportional to k_1vn^i where k_1 is a constant, v represents the number of bands, n is the number of pixels, and i ranges from 1 through 3. In contrast, most parametric classifiers execute in time approximately proportional to k_2vcn where k_2 is a constant, and c is the number of classes. It should also be noted that a clustering algorithm usually requires multiple passes through a data set, whereas pixel assignment requires only a single pass; hence, $k_1 > k_2$ for identical values of n. To summarize, one partitioning of a complete image would require execution time proportional to k_1vN^i where N is the number of pixels in the image. The hybrid approach would only require time proportional to $k_1vn_s^i + k_2vcN$ where n_s is the number of pixels in the extracted sample. In typical projects where $c << n_s << N$, the hybrid is faster. When an identical number of iterations are performed during the clustering stages of both the hybrid and normal unsupervised processes, the time required for clustering repetitions rapidly overshadows the time required for the solitary pixel assignment step, and the comparison effectively reduces to k_2N^i versus $k_1n_s^i$ for the normal and hybrid methods, respectively. Given the exponential increase in execution time with a linear increase in n and the usual condition that $N>>n_s$, the difference in execution times between the two methods can be significant.

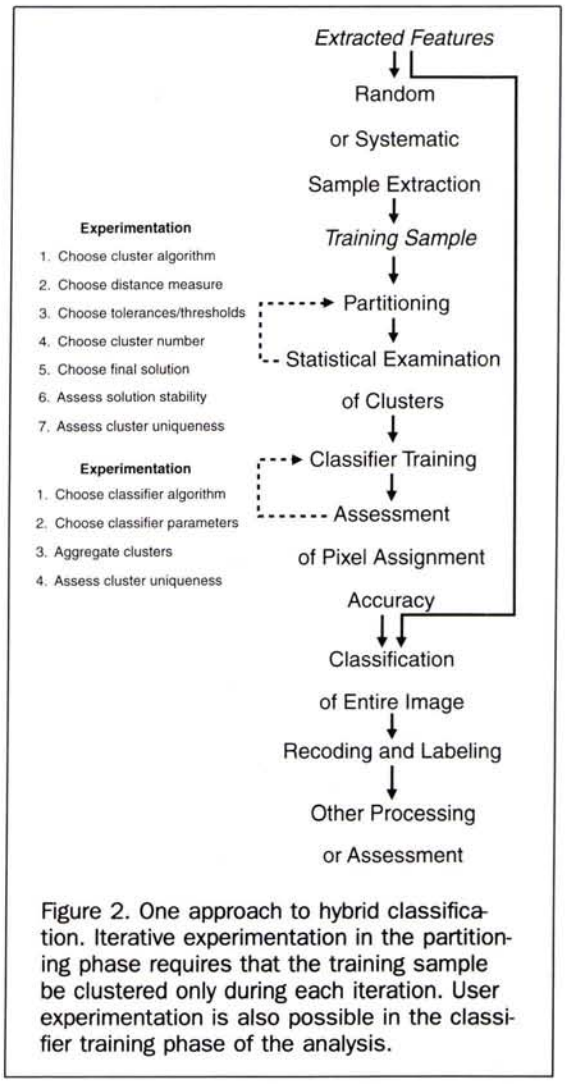

Figure 2. One approach to hybrid classification. Iterative experimentation in the partitioning phase requires that the training sample be clustered only during each iteration. User experimentation is also possible in the classifier training phase of the analysis.

This paper explores a simple question of hybrid classification – do nonparametric nearest-neighbor classifiers assign image pixels to spectral classes more accurately than the common parametric alternatives? While the question is simple, given the variety of Earth landscapes and different approaches to image processing, a general answer is impossible. Nevertheless, this paper presents the results of a project to compare the pixel assignment accuracy of six nearest-neighbor classifiers and four parametric classifiers across six small Landsat images. The results reported in this paper should encourage readers to examine nearest-neighbor classifiers as alternatives to parametric methods for hybrid classification, especially when training data sets are large and parametric classification accuracy is low.

In the next section, a brief history of nearest-neighbor classification will be presented so that this research can be placed in its historical context. A brief discussion of parametric classifiers will follow. This section will focus primarily on the statistical definitions of their discriminant functions, and their statistical behavior. Several nearest-neighbor classifiers will then be discussed. While a multi-

tude of nearest-neighbor classifiers can be formulated, the discussion will be limited to six fundamental variants. In the succeeding section the methodology of the comparative experiments will be presented. Content of the six images will be briefly described, the method used to generate the classifications will be outlined, and the statistical measures used to compare the accuracy of the classification will be reviewed. Results of the experiments will be discussed and a summary will be offered in conclusion.

Historical Background

Nearest-neighbor classification has been studied for at least four decades. Credit for the first formulation of nearest-neighbor rules is given to Fix and Hodges (1951). These authors developed the *k* nearest-neighbor rule as an outgrowth of their research to model multivariate density functions nonparametrically. In the context of pixel assignment, the *k* nearest-neighbor rule would assign an unlabeled pixel to the majority class represented by a prespecified number (i.e., *k*) of training pixels spectrally adjacent to it. Sixteen years later, Cover and Hart (1967) theoretically examined the first nearest-neighbor rule as a special case of this *k* nearest-neighbor rule. As the name suggests, a pixel being labeled using the first-nearest-neighbor rule would take the label of its closest spectral neighbor among the training pixels. In 1970, Patrick and Fischer proposed a new definition of neighborhood. Rather than specifying a number of nearest neighbors, they proposed to define a neighborhood as a hypersphere of fixed radius centered on the query point.

According to Hardin and Thomson (1992), pixel assignment based on spectral nearest neighbors has not been widely reported in remote sensing, primarily because the slow rate of pixel assignment precludes its practical use with large multiband images. However, those authors review a combination of data structures and algorithm advances which make nearest-neighbor classifiers practical alternatives to parametric classifiers from a computational perspective. While these classifiers may now be practical for remote sensing from a computing viewpoint, the accuracy of their pixel assignment has not been thoroughly examined. Ince (1987) compared a multiple neighbor-based rule to a maximum-likelihood classifier for a agricultural scene of Turkey and found it to be superior. More recently, Skidmore (1989) determined the first nearest-neighbor rule to be poor in discriminating between varieties of Australian eucalypt forest.

Parametric Methods

As outlined by Jensen (1986), parametric classifiers assume that the populations which provide training data are multivariate normal across all the selected features for each nominal class. The parameters typically utilized in parametric classification include feature mean vectors (μ), variance-covariance matrices (ζ), and prior probabilities (p) for each class (h) where $h = 1, \ldots, g$. According to Swain (1978), when given a measurement vector ($\mathbf{x}$) and the class conditional probability for a given class $p(\mathbf{x} \mid h)$, the maximum-likelihood decision rule demands we assign an unknown observation to class h if $(p_h)p(\mathbf{x} \mid h) \geq (p_i)p(\mathbf{x} \mid i)$ for all $i \neq h$. Some authors have applied the term "maximum-likelihood" to classifiers whose prior probabilities are unspecified. In these cases, the decision rule reduces to making the assignment to class h if $p(\mathbf{x} \mid h) \geq p(\mathbf{x} \mid i)$ for all $i \neq h$. While this paper explores parametric classifiers using both decision rules, the term "maximum-likelihood" is reserved for the rule requiring prior probabilities.

Mahalanobis' Distance Function

The Mahalanobis' distance function (MDF) is used for individual pixel assignment by calculating the Mahalanobis' distance (D^2) from the unlabeled pixel to each training group centroid and assigning the pixel to the group generating the smallest Mahalanobis' distance. Specifically, the distance from an unclassified pixel (j) to the centroid of group h can be expressed as

$$D^2_{jh} = (\mathbf{x}_j - \mu_h)^T \zeta^{-1} (\mathbf{x}_i - \mu_h),$$

where ζ is the pooled group variance-covariance matrix (or common covariance matrix). The Mahalanobis' distance measure can be thought of as a minimum-distance-to-means classifier which (1) adjusts the Euclidean distance between a pixel and a candidate class by the correlation between the feature dimensions and (2) removes the bias introduced by scale differences among the features.

The MDF assigns pixels to classes without regard to class prior probabilities and thus does not adjust decision boundaries to minimize an overall error rate. For optimum classification based on the MDF, the group covariance matrices must be equal, and the data in each group must be multivariate normal across all the features. As with other parametric classifiers, the consequences of violating these conditions is less severe when interclass variances are smaller, intraclass distances are greater, or the query pixel is closer to the correct class centroid.

The Linear Discriminant Function

In social statistics, the general workhorse for discrimination is Fisher's linear discriminant function (James, 1985). The formula for determining the linear discriminant function (LDF) distance between pixel j and group h can be expressed as

$$F_{jh} = \mu_h^T \zeta^{-1} \mathbf{x}_j - 1/2\ \mu_h^T \zeta^{-1} \mu_h + \ln (p_h).$$

Like the MDF, the LDF distance is calculated between the unlabeled pixel and each group. The pixel is then assigned the label of the group that generates the smallest discriminant score. When the feature measurements are sampled randomly from populations which are multivariate normal and have identical covariance matrices, the LDF classifier qualifies as a maximum-likelihood classifier.

Quadratic Discriminant Function

The previous two classifiers required that the group covariance matrices be equal in order to maximize classification accuracy. In contrast, this constraint can be relaxed when quadratic discriminant functions are used. As shown in the equation below, the individual group covariance matrices (ζ_h) rather than a pooled group covariance matrix are applied in the classification to achieve this independence. If we designate the determinant of the covariance matrix for group h as d_h, the equation for the quadratic discriminant function distance between pixel j and group h is given by

$$Q_{jh} = (\mathbf{x}_j - \mu_h)^T \zeta_h^{-1} (\mathbf{x}_j - \mu_h) + B_h,$$

where $B_h = \ln(d_h)$ for the quadratic discriminant function without priors (QDF) and $B_h = -\ln (p_h) + \ln(d_h)$ for the quadratic form with prior probabilities (QDP). Like the other classifiers, once the function calculations have been made for all groups ($h = 1, \ldots, g$), the pixel is assigned to the group which generates the smallest score. Like its linear counterpart, the quadratic discriminant function with priors can be

derived from Bayes' rule, and is a true maximum-likelihood classifier when the normality constraints are satisfied.

Nearest-Neighbor Methods

Nearest-neighbor methods are nonparametric – that is, no assumptions are made regarding the shape of each group's multivariate distribution. However, this does not imply that sampling method or sample size can be disregarded for the classifiers to operate efficiently. Neither does it mean that *a priori* information is not required for optimizing classifier accuracy.

Methods predicated on nearest neighbors have many potential advantages. First, because they are non-parametric, deviations from normality pose no hazard to the accuracy of the pixel assignment. Second, because the entire training set is used in the pixel assignment process, no information is lost by generalizing the data distribution through use of mean vectors or covariance matrices. Third, they are extremely flexible, and can be modified to support different neighborhood definitions and various heuristic means of class and distance weighting. Finally, some nearest-neighbor classifiers can be derived from Bayes' formula, and are maximum-likelihood classifiers when the training set is properly specified (see James, 1985).

Three components are common to all nearest-neighbor classifiers. In the context of image processing, the first is a definition of a query pixel's neighborhood. Typically, a neighborhood is defined either as (1) a prespecified number of training pixels closest to the query pixel's feature coordinates or as (2) a radius centered on the query pixel's feature coordinates. To limit the scope of this paper, only the first definition will be utilized. The second component is a search algorithm. For any given query pixel, this algorithm extracts the training pixels which satisfy the neighborhood definition. Finally, a decision criteria must be specified to label the query pixel predicated on the classes represented in the extracted neighborhood.

The *k*-Nearest-Neighbor Rule

As mentioned previously, Fix and Hodges (1951) developed the *k*-nearest-neighbor (*k*-NN) rule in an attempt to nonparametrically model multivariate density functions. In this rule, a neighborhood is defined as a fixed number of pixels (k) centered on the unlabeled pixel's coordinates ($\mathbf{x}$) in feature space . The class labels of these neighboring training pixels are examined, and the class label represented most frequently is given to the query pixel. In performing classification with the *k*-NN rule, the practitioner must specify the number of pixels to be utilized as a neighborhood, and the manner in which ties in the "voting" will be resolved. In Figure 3, the neighborhood for the query pixel is predefined as its seven nearest neighbors. The unlabeled pixel is assigned to Class 2, because it constitutes a majority of the seven pixels.

A substantial body of literature, collected and discussed by Dasarathy (1991), is devoted to an examination of the statistical characteristics of nearest-neighbor rules. Summarizing a portion of this collection, when the proportion of pixels in each training class is identical to the actual proportion of each class in the population, the *k*-NN rule is a maximum-likelihood classifier.

The First Nearest-Neighbor Rule

The simplest rule predicated on spectral neighborhoods is the first nearest-neighbor (FNN) rule (see Cover and Hart, 1967). In the context of pixel assignment, a first nearest neighborhood would be defined as the single training pixel with the shortest Euclidean distance[2] to the unlabeled pixel. The unlabeled pixel is assigned to the class of this nearest neighbor. This criterion is a special case of the *k*-NN rule discussed above, yet it is considered separately because of its great popularity in the social sciences and engineering. Like the *k*-NN rule, the FNN rule is also a maximum-likelihood classifier when the training class pixel proportions mirror the population proportions. However, the FNN rule always results in an unambiguous decision – because only one vote is taken to assign a query pixel to a class, no ties are possible.

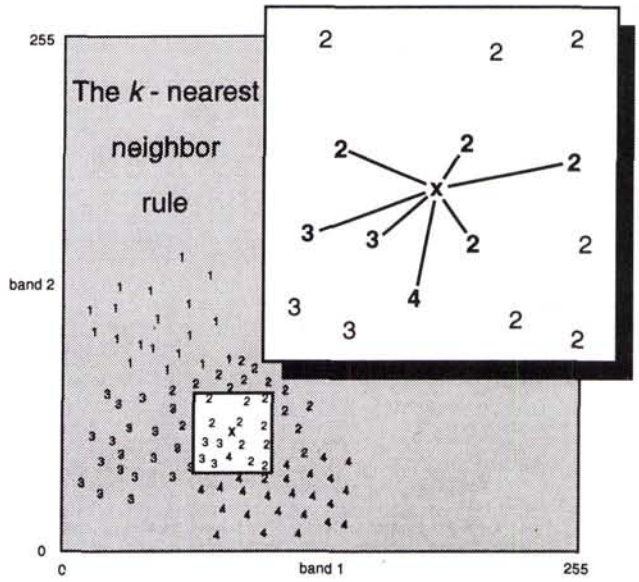

Figure 3. The multiple (k) nearest-neighbor rule.

The Distance-Weighted Nearest-Neighbor Rule

As a modification of the multiple nearest-neighbor rule, Dudani (1976) proposed that training samples closest to the query coordinates should cast votes which have greater weight than training samples more distant. In general, this rule can be termed the distance-weighted nearest-neighbor (DWN) rule. The DWN neighborhood is identical to the *k*-NN neighborhood, but the decision criterion is quite different – the unlabeled pixel is assigned to the class producing the highest aggregate weight across the prespecified (i.e., k) training pixels. To employ the DWN rule, the user must specify the distance function for weighting as well as the number of nearest neighbors to be polled. As discussed by Macleod *et al.* (1987), the results of any classification are very dependent on this weighting function. Consider the problem depicted in Figure 3. If the vote cast for the query pixel by each of its seven nearest neighbors was weighted by the inverse of its intervening Euclidean distance, the unlabeled pixel would again be assigned to Class 2.

[2]Some practitioners in other scientific fields utilize Mahalanobis' distance rather than the Euclidean distance, put the practice does not seem to be general in applications such as image processing where extremely large data sets require classification.

The Rank-Weighted Nearest-Neighbor Rule

When employing the DWN classifier, the absolute metric distances are used in some weighting scheme to predict the query pixel label. In contrast, the rank-weighted nearest-neighbor (RWN) rule requires that weights be assigned to each of the k nearest neighbors according to their relative (rather than absolute) distances from the query pixel. Like the DWN rule, neighbors closer to the query pixel would still be weighted more heavily than neighbors more distant, and different weighting schemes could change the classification results substantially. For example, in classifying the query pixel in Figure 3, the first neighbor might be weighted more heavily than the other neighbors. It might also have been decided that the combined votes of the second and third neighbors should be equal to the first neighbor's vote. The same general logic might be extended to the seventh neighbor. Using this logic, the query pixel in Figure 3 would again be placed in Class 2.

The Class-Weighted Nearest-Neighbor Rule

In projects where a different cost is associated with misassignment for each class, the k-nearest-neighbor decision criteria can be roughly adjusted to account for it. For example, we might assume that the goal of our hypothetical classification project (i.e., shown in Figure 3) is to create a satellite-derived map of possible wetlands to support a more intensive low altitude non-riparian wetland detection project. We may have also determined that Classes 3 and 4 represented non-riparian and riparian wetland, respectively, while Classes 1 and 2 were land-cover types of lesser interest. In order to avoid missing any possible non-riparian wetland, it was decided to assign Classes 3 and 4 weights of 10 and 7, respectively, while smaller weights were assigned to Classes 1 and 2. In devising this weighting scheme, we were willing to risk overestimating wetland extent in order to facilitate the project objectives. Given this scenario, the class-weighted nearest-neighbor rule (CWN) assigns the unlabeled pixel to Class 3, even though Class 2 supplies the majority of the seven neighboring pixels.

In cases where project goals do not suggest a class weighting scheme, the CWN rule can still be useful. In an initial classification, the class weights are all set to 1.0, effectively reducing the CWN rule to the k-NN rule. The weights are then iteratively adjusted by the operator to compensate for misassignment between various classes.

Bayesian Nearest-Neighbor Rule

As mentioned above, the k-NN rule is a maximum-likelihood classifier when the training sample class proportions match the class proportions in the population to be classified. However, when the training set violates this condition, the practitioner can either discard training data until the condition is satisfied, or adjust the k-NN voting scheme by utilizing known prior probabilities.

It is common knowledge that Bayes' theorem takes the form

$$p(h \mid \mathbf{x}_j) = \frac{p(\mathbf{x}_j \mid h)(p_h)}{\sum_{i=1}^{g} p(\mathbf{x}_j \mid i)\ p(i)}.$$

Assume a fixed number of pixels for application of the k nearest-neighbor rule (k). As reviewed by James (1985), if K_h is the number of nearest neighbors in class h among this prespecified number k, an estimate of $p(h \mid \mathbf{x}_j)$ derived from Bayes' rule can be calculated by

$$p(h \mid \mathbf{x}_j) = \frac{K_h\ p_h}{\sum_{i=1}^{g} K_i p_i}.$$

In classification, the unlabeled pixel would be assigned to the class producing the highest probability $p(h \mid \mathbf{x}_j)$. The advantage of this Bayesian nearest-neighbor (BNN) rule over the other neighbor-based rules is that the training sets do not have to be gathered in proportions representative of their respective populations in order to retain the desired Bayesian behavior. Following the hypothetical wetlands example introduced above, the prior probabilities of 0.20, 0.40, 0.10, and 0.30 might reflect our estimates of the relative area covered by land-cover classes one through four, respectively. We might also have training pixel total counts of 19, 25, 23, and 25, respectively, for the same four classes. The training pixel would be assigned to Class 2, because the *a posterior* probability estimate (0.76) calculated for Class 2 exceeds that of the competing classes.

Methodology

The Imagery

The hybrid classification tests described below were performed on six EOSAT images. These *Sample Thematic Mapper Floppy Disk Data Products* were selected because they provided a variety of land-cover combinations, were inexpensive, and would remain readily available to other researchers. Adapting some segments from descriptions found in the *Eosat Landsat Products and Services* guide, each of the images will be characterized below.

Latour

The Latour image embraces an area of Mississippi River bottomland near Helena, Arkansas, and is typical of a Mississippi delta region where cotton, soybeans, and winter wheat dominate the agricultural landscape. The river is the central feature of the image, and the contrast between the silted river and clear oxbow lakes on the image is striking. Given the winter date (13 January 1983), most variation in the scene's agricultural components can be attributed to soil moisture, weed growth, and the vigor of some winter and volunteer crops. In addition, large blocks of forested bottomland on the image appear near the river, and smaller blocks are interspersed among the agricultural fields. Meander scars and sand bars with high albedo are also conspicuous.

Morro Bay

This image, acquired by Landsat 5 on 19 November 1984, shows Morro Bay, California and approximately 10,000 hectares of the nearby Pacific Coast Range. Vegetation at the lower mountain elevations can largely be described as chaparral. Hardwoods appear at higher elevations, and brightly reflective areas of sparse vegetation are visible at all elevations. Some agricultural fields at various stages of production are visible in the valleys, but these are spectrally similar to adjacent golf courses and suburban lawns in the more populated areas. Infrared-bright wetlands are also clearly discernible in the image.

San Joaquin

This Landsat 5 image, acquired on 5 September 1986, covers a portion of high-value agricultural land in the San Joaquin

Valley southwest of Bakersfield, California. A careful visual inspection of the landscape betrays over a dozen different classes, including secondary roads, Interstate 5, orchards, ponds, and oil fields, but the largest component of spectral variability is due to the variety of crops in various phases of production. Bare soil with different moisture and color characteristics also contributes to the spectral complexity of the scene.

Little Colorado
This 24 August 1985 Landsat 5 image is centered on the confluence of the Colorado and Little Colorado rivers in the Grand Canyon of Arizona. Given the paucity of vegetation and late summer date, most of the variation in the image can be traced to surficial geology, and geologic exposures in the canyon itself. Throughout the image, shadowing is strong, and the dissection of the region is starkly apparent.

New Orleans
This image, acquired by Landsat 5 on 24 March 1985, covers the northern limits of New Orleans, Louisiana, along its shoreline with Lake Pontchartrain. Very complex formations of wetland vegetation are particularly conspicuous between the Lake and a small portion of the silt laden Mississippi River. The thermal band of this image is particularly useful for demarcating the boundary between older urban areas and the "cooler" surrounding environs.

Black Hills
This scene is a section of the northern Black Hills adjacent to Lead, South Dakota (which appears in the southwest corner of the image). As the town name implies, evidence of lead mining is apparent, including stripped areas devoid of vegetation, and large areas of mine refuse in various stages of reclamation or secondary regrowth. Taken in mid May of 1985 by Landsat 5, some small-grains agriculture (clearly dryland rather than irrigated) is visible in the northwestern quarter of the image. This is mixed with rangeland which forms some patches within the main mountainous areas. Where mining has spared the higher mountain landscape, ponderosa pine forests are visible.

Training and Testing
The classification methodology was identical for each of the six Landsat images. First, a random sample of 15,000 row and column coordinate pairs were generated, and the corresponding brightness values for all seven TM bands were extracted from the image. This sample was then partitioned using the SPSSPC quick cluster routine where a 20-cluster solution was requested. The choice of 20 cluster classes was arbitrarily selected as an upper limit which could be managed efficiently in the accuracy assessment procedures described below.

The SPSSPC quick-cluster routine was chosen because it is readily available to researchers and performs partitioning in a manner similar to many clustering routines in commercial image processing software. As described in Norusis (1988), this cluster routine is predicated on the sorting of nearest centroids. In the context of this research, an initial number of clusters (M) is first specified. A set of M well spaced pixels in the data set are selected as initial cluster centers. As pixels are sequentially processed and assigned to the nearest cluster center, any pixel may supplant a particular cluster center if the interposing Euclidean distance is larger than the distance between any pair of cluster centers. Furthermore, any pixel will replace a current cluster center if its Euclidean distance to the center is larger than the shortest distance between the center and all other current cluster centers. Once an initial pass through the data is complete, a classification cluster centroid is calculated for each cluster class from all its constituent pixels. A second pass through the data then reassigns all the pixels to the closest classification cluster centroid.

When this clustering of the 15,000 pixels was completed, all clusters with less than 60 pixels were discarded from the analysis, resulting in a final cluster solution of between 8 and 18 clusters, depending on the image. The limit of 60 pixels was chosen to ensure at least 30 pixels per cluster would be used for both training and testing. Clusters were not removed from the analysis for any other reason.

Using two methods, pixels within the retained clusters were divided into training and verification sets. In the first case, the pixels for each cluster were evenly split between the two sets. This ensured that the proportion of pixels resident in each class of the verification set was equivalent to pixel proportions in the training set, and satisfied the optimum condition for classification with the nearest-neighbor classifiers. The large size of the "full" training set also guaranteed accurate summary statistics for training the parametric classifiers. In the second case, there was a deliberate attempt to severely violate the optimum condition for nearest-neighbor classification – class proportions in the sample matching those in the population. For a given image, the cluster class with the lowest pixel count was identified, and this same number of pixels was extracted from the remaining clusters in the full training set to created a reduced training set. When this subsampling was performed on each of the six full training sets, it produced subsets containing only between 30 and 62 pixels per cluster.

Splitting the data set into training and verification sets was a variation from the normal hybrid classification process. In normal hybrid classification, once the clustering of the sample is performed, statistics on each cluster class are calculated and used to train a classifier. The classifier is then used to place every image pixel into a predicted spectral class. In this simulation, the training sets assume the role of systematic or random samples, and the verification set is analogous to the entire image to be classified. Sampling 15,000 pixels and clustering them into initial spectral classes was performed because it permitted the pixel assignment experiments to be conducted using a pixel population with a known underlying spectral structure. In practical hybrid classification, this structure is unknown. However, with this known spectral structure in the simulation, a definition of "correct spectral class" could be formulated for accuracy assessment of the pixel assignment step.

No other preprocessing, band selection, or standardization was performed. The ten classifiers described above were all trained separately using the full and reduced training sets. They were then tested using the verification set to determine how well the various classifiers could assign pixels to their correct spectral class. A pixel's correct spectral class was defined as the cluster class it was placed into by the original clustering of the 15,000 pixels. Some other notes about the classification strategy are listed below.

- For classifiers using prior probabilities, the actual class proportions in the verification set were used.
- Seven neighbors were specified for all the classifiers requiring multiple nearest neighbors.

TABLE 1. TABLEWISE KAPPA VALUES FOR THE SIX TEST IMAGES. THESE CLASSIFICATION EXPERIMENTS UTILIZED THE FULL (UNREDUCED) TRAINING DATA SETS. THE HIGHEST RAW KAPPA VALUE FOR EACH IMAGE IS SHOWN BY AN ASTERISK. THE BOLD VALUES IN THE TABLE REPRESENT KAPPA VALUES STATISTICALLY EQUIVALENT TO THE HIGHEST KAPPA.

	Image					
Classifer	Latour	Morro Bay	San Joaquin	Little Colorado	New Orleans	Black Hills
Mahalanobis' Distance Function (MDF)	0.869	0.880	0.800	0.868	0.838	0.873
Linear Discriminant Function (LDF)	0.906	0.920	0.870	0.895	0.886	0.906
Quadratic Discriminant Function (QDF)	0.896	0.872	0.814	0.849	0.840	0.847
Quadratic Discriminant Function (QDP)	0.891	0.896	0.843	0.878	0.866	0.881
k - Nearest Neighbor (*k*-NN)	**0.959**	**0.960**	**0.932**	**0.945**	**0.960**	**0.963**
First Nearest Neighbor (FNN)	0.948	0.941	0.919	0.923	0.946	0.945
Distance Weighted Neighbor (DWN)	**0.961***	**0.961***	**0.937***	**0.946***	**0.961***	**0.964***
Class Weighted Neighbor (CWN)	**0.958**	0.945	0.924	**0.944**	**0.958**	0.950
Bayesian Nearest Neighbor (BNN)	**0.959**	**0.960**	**0.931**	**0.944**	**0.960**	**0.963**
Rank Weighted Neighbor (RWN)	**0.956**	**0.953**	**0.932**	**0.936**	**0.956**	**0.960**

- For the RWN classifier, the rank weights assigned for neighbors 1 through 7 were 64, 32, 16, 8, 4, 2, and 1, respectively.
- No external project criteria were specified for use with the CWN classifier. Therefore, weights for the CWN classifier were chosen after examining the results of the FNN classification in order to minimize misclassification. Classes with high exclusion error rates were weighted heavily, and, conversely, classes with high inclusion error rates were weighted lightly. Only one attempt at this "fine tuning" was allowed.
- Each nearest neighboring pixel used in the DWN classification was weighted according to the inverse of its squared distance from the unlabeled pixel.
- In all of the multiple nearest-neighbor classifiers, ties were handled by assigning the pixel randomly to one of the tied classes.

Accuracy Assessment

After the pixel assignment was complete, agreement between the actual clusters and the predicted clusters was presented in confusion matrix form for accuracy assessment. The agreement between the actual and predicted classes was measured using Cohen's kappa. For a single image, the results of the several classifiers were compared statistically on a pairwise basis using the two-tailed Z test for kappa described by Congalton et al. (1983). An alpha level of 0.01 was chosen as the acceptable level of significance – any Z with an absolute value greater than 2.576 was considered to represent a significantly different pair of kappa accuracy rates.

Rather than showing the confusion matrices for each classification (i.e., 120 matrices), the single parametric and nonparametric classifiers which produced the highest raw kappa values were alone retained for further analysis. The confusion matrices for these "winners" were first standardized (see Congalton *et al.*, 1983). The main diagonals from these standardized matrices were next extracted and tabled in column form. Because this standardization accounted for differences in predicted group size, the best parametric and nonparametric classification rates for each image could be compared on a classwise basis.

A tablewise accuracy figure was also calculated by computing the mean of the diagonal elements in the standardized table. This figure, compared with the computed median of the standardized diagonal elements, indicated whether high (or low) overall classification accuracies were due to a few well (or poorly) assigned classes with large numbers of pixels, or whether there was a consistent accuracy rate across all the classes.

Results

The classification results for the six images are summarized in Tables 1 and 2. Table 1 shows the results using the full training set, while Table 2 shows the results from the reduced training set. The highest raw kappa for each image is indicated by an asterisk. For each image, the bold values in the column are statistically equivalent to the highest (aster-

TABLE 2. TABLEWISE KAPPA VALUES FOR THE SIX TEST IMAGES. THESE CLASSIFICATION EXPERIMENTS UTILIZED THE REDUCED TRAINING DATA SETS. THE HIGHEST RAW KAPPA VALUE FOR EACH IMAGE IS SHOWN BY AN ASTERISK. THE BOLD VALUES IN THE TABLE REPRESENT KAPPA VALUES STATISTICALLY EQUIVALENT TO THE HIGHEST KAPPA.

	Image					
Classifer	Latour	Morro Bay	San Joaquin	Little Colorado	New Orleans	Black Hills
Mahalanobis' Distance Function (MDF)	0.850	0.852	0.795	0.817	0.831	0.857
Linear Discriminant Function (LDF)	**0.894***	**0.885**	**0.865***	**0.861***	**0.870***	**0.884***
Quadratic Discriminant Function (QDF)	0.813	0.768	0.797	0.761	0.774	0.757
Quadratic Discriminant Function (QDP)	0.838	0.785	0.820	0.793	0.796	0.782
k-Nearest Neighbor (*k*-NN)	0.864	**0.877**	0.845	0.832	0.823	0.867
First Nearest Neighbor (FNN)	**0.879**	**0.877**	0.846	**0.852**	**0.857**	0.868
Distance Weighted Neighbor (DWN)	0.869	**0.884**	**0.855**	0.837	0.834	**0.875**
Class Weighted Neighbor (CWN)	0.846	0.845	0.839	0.805	0.820	**0.872**
Bayesian Nearest Neighbor (BNN)	0.863	0.874	0.846	0.830	0.823	**0.873**
Rank Weighted Neighbor (RWN)	**0.882**	**0.890***	**0.859**	**0.846**	0.853	**0.873**

TABLE 3. A SUMMARY OF THE CLASSIFICATION EXPERIMENTS FOR THE SIX IMAGES

Result	Image					
	Latour	Morro Bay	San Joaquin	Little Colorado	New Orleans	Black Hills
Number of classes retained	11	15	18	12	10	8
Total number of test pixels	7469	7451	7479	7425	7458	7389
Mean number of test pixels per class	679	470	416	619	746	924
Number of test pixels in smallest class[1]	49	30	62	57	31	31
Number of test pixels in largest class	2354	1762	1258	2725	2339	2519
			Using full training set			
Best parametric classifier	LDF	LDF	LDF	LDF	LDF	LDF
Kappa of best parametric classifier	0.91	0.92	0.86	0.90	0.89	0.91
Mean/Median standardized accuracy	90/93	89/89	86/88	91/91	89/89	89/89
Best nonparametric classifier[2]	DWN	DWN	DWN	DWN	DWN	DWN
Kappa of best nonparametric classifier	0.96	0.96	0.94	0.95	0.96	0.97
Mean/Median standardized accuracy	94/94	91/92	92/92	94/93	95/94	96/95
Statistical difference between best parametric and best nonparametric kappas?	Yes	Yes	Yes	Yes	Yes	Yes
			Using reduced training set			
Best parametric classifier	LDF	LDF	LDF	LDF	LDF	LDF
Kappa of best parametric classifier	0.89	0.89	0.87	0.86	0.87	0.89
Mean/Median standardized accuracy	90/91	87/85	87/88	89/89	88/90	88/90
Best nonparametric classifier[2]	RWN	RWN	RWN	FNN	FNN	DWN
Kappa of best nonparametric classifier	0.88	0.89	0.86	0.85	0.86	0.88
Mean/Median standardized accuracy	92/93	88/88	91/91	90/90	90/90	91/91
Statistical difference between best parametric and best nonparametric kappas?	No	No	No	No	No	No

[1]These cells also indicate the number of pixels in every class of the reduced training sets.
[2]As measured by the highest raw kappa value. As mentioned in the text, some other kappa values were statistically equivalent.

isked) kappa. A summary of the best classifications is recited in Table 3.

Full Training Set Results

When the full training set was used (Table 1), the neighbor-based classifiers produced the highest accuracy for all six images. With exception of the FNN classification of all the images, and the CWN classification of three images, all the neighbor-based classifiers produced significantly higher accuracies than all the parametric classifiers – never was any parametric classifier statistically superior to any nonparametric classifier. However, disregarding the nine exceptions already cited, the accuracy differences produced for each image among the neighbor-based classifiers alone were insignificant (see the bold values in Table 1). As shown in Table 3, the DWN classifier consistently produced kappas with the highest raw value.

Among the parametrics, the LDF rule invariably yielded the highest kappa values (see Table 3). In all cases, this superior accuracy was significantly different from the other parametric classifiers. With one exception, the QDP classifier came in second place among the parametric classifiers.

Reduced Training Set Results

The results achieved with the reduced training set (see Table 2) were quite different from those produced with the full training set. With the exception of Morro Bay, the LDF classifier produced higher raw kappa values than any of the other nine methods. Nevertheless, for all five of these images, there were at least two nonparametric classifiers where this difference proved statistically insignificant. For each of these five images, the RWN classifier produced accuracies statistically equivalent to the parametric LDF classifier. In the case of Morro Bay, the RWN classification accuracy was higher than the accuracy achieved by the LDF classifier, but the difference was not significant.

When considered alone as a group, the results of the nonparametric classifications using the reduced training set were much more inconsistent than the full training set results. Unexpectedly, the BNN classifier, designed to take into account training pixel proportions not reflecting the population, was equivalent to the highest kappa for only one image (Black Hills). The FNN classifier was also expected to perform well under full training set conditions but poorly under reduced training set conditions. The opposite was generally true – in four of the six images, it produced accuracies equivalent to the highest accuracies when the reduced training set was used, and never produced accuracies equivalent to the highest in the full training set.

Although the relative performance of the LDF classifier, when compared to the other parametrics, was undeviating under both full and reduced training set conditions, the QDP classifier was replaced by the MDF classifier as the second place parametric under reduced training set conditions. The San Joaquin image was a single exception – the QDP classifier remained superior to the MDF method, but was still inferior to the LDF rule.

Other Observations

The informational classes corresponding to the spectral classes produced in the clustering logically follow from the short descriptions of each image offered earlier in the paper. A de-

tailed narrative of the information classes is not included here, but is available from the author. In several instances, multiple spectral classes corresponded to a single informational class. Likewise, some spectral classes were mixes of various informational classes. To further shorten this paper, the results of the matrix normalization were also removed. Summarizing across the 12 experiments (six images by two training sets), there were a total of 296 main diagonal cells in 24 normalized matrices representing the best parametric and nonparametric classifiers for each image. The average accuracy difference among these 296 cells was approximately 3 percent (standardized), and in only two cells did the difference exceed 10 percent (standardized). No pattern of classification difference was apparent – the classification behavior difference between the best parametric and nonparametrics was independent of the individual classes involved. Using the reduced training or the full training set was inconsequential. Furthermore, the closeness of the mean and median standardized accuracy figures (see Table 3) indicated that the high accuracies achieved were due to effective pixel assignment in all the categories rather than just a few of the largest classes. No obvious difference in this relationship was revealed when the reduced training set was employed.

The loss of accuracy experienced when using the reduced training set instead of the complete training was particularly noticeable among the nonparametric classifiers. While the median decrease in accuracy between the full and reduced training sets was only 0.02 for the LDF classifier (the best parametric), the difference was 0.09 for the DWN classifier (the best nonparametric). The QDP classifier was also quite sensitive to the training set reduction, with an average difference of 0.07. It appears that pooling the group covariance matrices for use by the MDF and LDF classifiers renders some robustness against the effects of small training pixel counts. In contrast, while the QDP rule relaxes the prerequisite of having equal class covariance matrices, this advantage is offset by this loss of robustness afforded by the pooling process.

Summary

As summarized in Table 3, when the training data sets are large, and contain the same class proportions as the population to be classified, the neighbor-based classifiers, as a group, are statistically superior to the best parametric classifiers. The DWN classifier produced the highest accuracy values, but there were several other neighbor-based classifiers which were equivalent. In all the experiments, the LDF classifier was the superior parametric classifier. The nonparametric classifiers suffered more accuracy loss than the parametric classifiers when the reduced data set was used; however, there was at least one nonparametric classifier that still performed as well as the best parametric. The RWN, FNN, and DWN classifiers were the superior nonparametric classifiers under reduced training set conditions.

Given hybrid classification problems similar to the ones simulated in this project, what are the practical implications of these results? First, whether or not the training class proportions are equivalent to the true prior probabilities, the maximum-likelihood rule utilizing a linear discriminant function is probably a good classifier choice only if no neighbor-based classifiers are available. Second, if the training class proportions are severely different from the true prior probabilities, there is likely no advantage to be gained by using a neighbor-based classifier over the linear discriminant function – but neither will there be an accuracy penalty if the correct neighbor-based classifier is selected. Third, whether or not training class proportions are equivalent to the true prior probabilities, a distance-weighted neighbor classifier may be expected to give excellent results, and if the training class proportions are equivalent, DWN accuracy may be superior to any of the tested parametric alternatives. Fourth, when training class proportions are equivalent to their population counterparts, several of the neighbor based classifiers would be expected to perform equally well. Finally, theoretical advantages or disadvantages of certain classifiers such as the QDP, BNN, and FNN may disappear when tested under practical conditions, and the analyst is well advised to experiment with several classifier options.

While the purpose of this research was not to compare the parametric classifiers with each other, a single comment will be offered. As mentioned previously, both the LDF and the QDP rules can be derived from Bayes' theorem, and are maximum-likelihood classifiers. In these experiments, the LDF classifier was superior in every instance to its quadratic counterpart. Practitioners utilizing commercial remote sensing software may want to determine which of the two maximum-likelihood classifiers are incorporated into their software, and encourage vendors to implement the linear classifier if not already available.

The results of this study should also encourage researchers utilizing supervised classification to study the benefit of nearest-neighbor classifiers in their pixel assignment tasks. As accurately portrayed by a reviewer, certainly the task of placing image pixels in *a priori* information classes is more challenging than placing image pixels in spectral classes, and the potential for improving classification accuracies with nearest-neighbor methods should not be ignored. This is clearly an area of study which warrants further research and comparative experiments.

Acknowledgments

The author thanks Brandon S. Plewe, graduate student of SUNY (Buffalo) Department of Geography for his skillful computer processing of the data sets. The author also thanks the Intergraph Corporation of Huntsville, Alabama for their support of this research. The comments of the reviewers and editors are also gratefully acknowledged.

References

Campbell, J.B., 1987. Introduction to Remote Sensing, Guilford Press, New York, 551 p.

Congalton, R.G., R.G. Oderwald, and R.A. Mead, 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques, *Photogrammetric Engineering & Remote Sensing*, 49 (12):1671-1678.

Cover, T.M., and P.E. Hart, 1967. Nearest-neighbor pattern classification, *IEEE Transactions on Information Theory*, 13:21-27.

Dasarathy, B.V., 1990. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques* (B.V. Dasarathy, editor), IEEE Computer Society Press, Los Alamitos, California, 447 p.

Dudani, S.A., 1976. The distance-weighted k-nearest neighbor rule, *IEEE Transactions on Systems, Man, and Cybernetics*, 6(4):325-327.

Fix, E., and J.L. Hodges, 1951. *Discriminatory Analysis: Nonparametric Discrimination*, Project 21-49-004, Report No. 4, USAF School of Aviation Medicine, Randolph Field, Texas.

Hardin, P.J., and C.N. Thomson, 1992. Fast nearest neighbor classification methods for multispectral imagery, *The Professional Geographer*, 44(2):191-201.

Ince, F., 1987. Maximum likelihood classification: optimal or problematic? A comparison with nearest neighbor classification, *International Journal of Remote Sensing*, 8(12):1829-1838.

James, M., 1985. Classification Algorithms, Wiley, New York, 209 p.

Jensen, J.R., 1986. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall, Englewood Cliffs, New Jersey, 379 p.

Macleod, J.E.S, A. Luk, and D.M. Titterington, 1987. A re-examination of the distance weighted k-nearest neighbor classification rule, *IEEE Transactions on Systems, Man, and Cybernetics*, 17(4):689-696.

Norusis, M.J., 1988. *SPSS/PC+ Advanced Statistics V2.0*, SPSS Inc., Chicago, 505 p.

Patrick, E.A., and F.P. Fischer III, 1970. A generalized k-nearest neighbor rule, *Journal of Information and Control*, 16(2):128-152.

Skidmore, A.K., 1989. An expert system classifies eucalypt forest types using Thematic Mapper data and a digital terrain model, *Photogrammetric Engineering & Remote Sensing*, 55(10):1449-1464.

Swain, P.H., 1978. Fundamentals of pattern recognition in remote sensing, Chapter 3 in *Remote Sensing: The Quantitative Approach* (P.H. Swain and S.M. Davis, editors), McGraw-Hill, New York, 396 p.

(Received 02 September 1992; revised and accepted 24 March 1993; revised 02 June 1993)

Perry J. Hardin
Perry Hardin is an Assistant Professor in the Geography Department at Brigham Young University and directs the BYU Laboratory for Geographic Information Analysis. Holding a Ph.D. in Geography (Remote Sensing) from the University of Utah, he is an investigator on several projects involving the integration of remote sensing and GIS for resourch management. His research interests include biogeography, image processing theory, climatology, and habitat modeling.

Call-for-Papers

ASA-CSSA-SSSA Bouyoucos Conference

Mission Inn, Riverside, CA; May 1-3, 1995

CONFERENCE TITLE

Application of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone

CO-SPONSORS

AGU, ASPRS, AWRA, GIS World

Oral presentations, poster papers and computer demonstrations are invited that describe emerging scientific technologies and applications of GIS used in modeling the movement of non-point source pollutants in the vadose zone and in groundwater. Papers involving spatial analysis & geostatistical applications, remote sensing techniques, GIS applications, regional-scale transport models, current methodologies of coupling GIS to solute transport models, uncertainty analysis, and regional-scale case studies pertaining to non-point source pollutants in the vadose zone are being solicited. Contributors should provide a presentation title proposal and an abstract of 200 words or less by **February 15, 1995**.

CONFERENCE OBJECTIVE

Explore the positive and negative aspects of the current use of GIS and other emerging technologies as tools for modeling the transport of non-point source pollutants (e.g., salts, fertilizers, pesticides, trace elements, etc.) through the vadose zone and into the groundwater. The goal of the conference is to evaluate the viability of using this multidisciplinary approach; to promote interest in this newly developing area of applied research; and to stimulate interaction between GIS & remote sensing specialists, solute transport modelers, and geostatisticians.

SCHEDULED KEYNOTE SPEAKERS

Policy-maker: Congressman George Brown, Jr. (Chairman of the House Science, Space and Technology Committee); GIS: Peter A. Burrough, Michael F. Goodchild, Jack Dangermond; Solute Transport Modeling: David R. Maidment, William A. Jury, Rien van Genuchten; Geostatistics: Andre' G. Journel; Scale Dependency: R. J. Wagenet.

Contact: Dennis L. Corwin, USDA-ARS, U.S. Salinity Laboratory, 4500 Glenwood Drive, Riverside, CA 92501. Tel.: 909-369-4819; Fax: 909-369-4818.