Regionalized Classification and its Application to the Dakota Aquifer, Hodgeman County, Kansas

Ricardo A. Olea and Geoffrey C. Bohling

Report to the Dakota Aquifer Program

Kansas Geological Survey
Open File Report 96-22
June 1996

Regionalized Classification and its Application to the Dakota Aquifer, Hodgeman County, Kansas

Ricardo A. Olea and Geoffrey C. Bohling

ABSTRACT

Regionalized classification is a numerical technique for the partition of a portion of the planet into volumes that are as internally homogeneous as possible and as different as possible from each other. The method works at its best when there are measurements of the same attributes at all control points evenly scattered around the area of interest.

Application of the procedure employing four cumulative thickness measurements in 231 wells in south-central Hodgeman county leads to the following findings concerning the heterogeneity of the amount of shale in the Dakota aquifer:

1. Delineation of thick basal sand deposits in the Dakota Formation suggest sedimentation along fluvial system draining the area to the northwest.
2. A thicker than normal and elongated accumulation of sand suggests the existence of a WNW channel during Cheyenne time.
3. Most favorable deposition of sands in mid-Dakota Formation was erratic and concentrated in the flanks of flood plains.
4. There is an extensive northwest-southeast band where deposition of sands was never favorable.
5. Continuous favorable condition throughout the entire genesis of the Dakota aquifer is haphazard and sporadic.

All findings are in agreement with previous general notions about the geology of the aquifer in Hodgeman county in particular and the state in general.

INTRODUCTION

An important task for earth scientists is to solve inverse problems by measuring regionalized attributes that then can be used to postulate natural processes that may have generated the phenomena that one can observe today. Considering that natural processes are fairly complex and that samplings are rarely large enough, simplification of reality imposing a man-made order commonly play an important part in inverse modeling.

Classification of specimens into groups is one of these simplifying models. In our context we will be dealing with sites instead of specimens or objects. In a regionalized classification the sampling domain is a multidimensional geographic space, usually a portion of planet Earth with dimension three. If one surveys a coregionalization by making several measurements per site, the survey defines another multidimensional space, a space with as many dimensions as the attributes considered in the sampling, commonly four or larger. In such a space one can depict each site as a point whose coordinates are the value of the attributes at such site. The assumption of regionalized classification is that the points in the attribute space are not uniformly and regularly distributed. Instead, as illustrated in Figure 1, there are zones with few or no points at all and other places with high concentrations of points. Notice that it is not possible to model such a relationship through regression.

The aim of regionalized classification is to explore the assumption that observations form distinct clusters in atribute space and to take advantage of it in case one finds grounds to support it. If that is the case, regionalized classification breaks the observations into groups that are as internally homogeneous as possible and as different as possible among them. Proximity in attribute space does not guarantee proximity in geographical space, yet in practice one finds that observations in the same attribute space cluster tend to stay close in the geo-

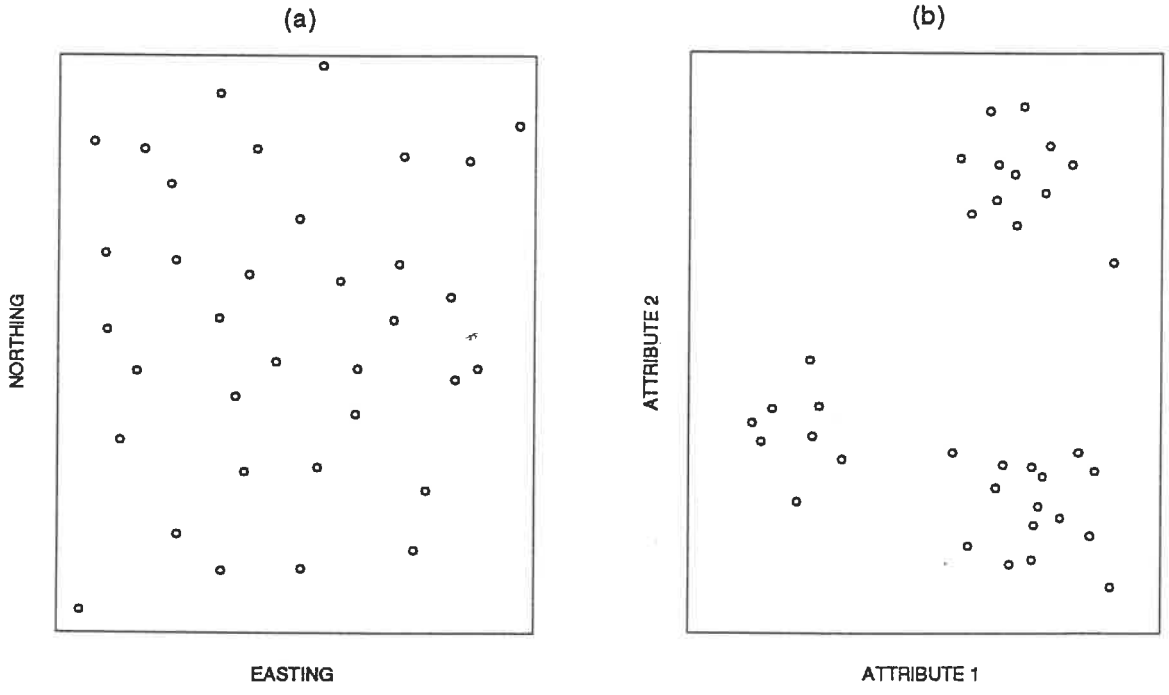


Figure 1. Idealized coregionalized sampling: (a) location of sampling sites in a two-dimensional space in a survey comprising measurement of two attributes per site; (b) mapping of the observations in the attribute space showing three clusters.

graphical space in various degrees. If one assigns contrasting colors to each group, the resulting group map in the geographic space resembles a geologic map, which is a type of regionalized classification map.

Another way to view regionalized classification is as a tool to condense maps. One can map all attributes of a coregionalization separately and inspect them to find relationships among features in the different maps. The task is simplified by the use of a light table, but becomes progressively difficult as the number of attributes in the survey increases. Regionalized classification is a convenient replacement of the light table to produce a single objective and accurate summary map, regardless of the number of attributes.

Conceptually nothing is new in regionalized classification. What is novel is the joint application of several standard techniques. Harff and Davis (1990) published the first complete formulation of the method combining geostatistical elements with some inspirational ideas of soviet geologists (Voronin, 1967; Rodionov, 1981; Kogan, 1986).

TYPIFICATION

In a nutshell, regionalized classification is the probabilistic assigment of sites to groups employing discriminant analysis. Discriminant analysis requires a training set that in regionalized classification ordinarily is prepared using cluster analysis. Preparation of maps may require the interpolation of group probabilities that may be done by kriging.

In classical applications of discriminant analysis, one employs a training set including direct indicators of group membership. Medical science provides good examples. In certain instances, a group of sick patients is carefully examined to prepare a clinical record relating internal conditions to their external symptoms and laboratory analyses, a costly and lengthy classification only possible after performing surgery or autopsy. In this context, based only on external symptoms and laboratory analyses, the task of discriminant analysis is to determine the probability that other patients suffer the internal conditions considered in the training set.

In earth sciences what prevails is the equivalent of knowing only the external symptoms and never knowing the group assignments. Even worse, the nature and number of the groups is seldom known. Determining the groups is as important in such studies as the assignments themselves. In such a situation one replaces the training set assignments by a typification using another method able to decide both the number and characteristics of the groups. The best alternative is cluster analysis, which aside from any prior information not contained in the data, solves the following problem: given a coregionalization sampling of size n comprising p attributes, is there any evidence for clustering the sites into groups scattered around centroids, as against the alternative hypothesis that they are an unstructured coregionalization?

There are several ways to define and find the clusters depending partly on the metric used to determine proximity among the sites in the p dimensional

attribute space, the Euclidean distance being the most common and the only one considered here. The geographical distance is completely ignored both in cluster analysis and in discriminant analysis and is only considered at the final mapping stage of regionalized classification.

There have been many studies comparing various methods of cluster analysis using artificial data sets containing known clusters produced by Monte Carlo methods. In most of these studies the Ward's minimum variance method has been the one with the best overall performance on reproducing the known clusters (SAS, 1990, p. 56).

WARD'S METHOD

Given n sites, this heuristic method progresses by reducing the number of clusters one at a time starting from the trivial extreme case of one cluster per site and ending at the other trivial extreme case in which one cluster comprises all the sites. At each cluster reduction, the method merges the two clusters resulting in the smallest increase in the total sum of squares of the distances of each point to its cluster centroid. Sites clustered at a previous clustering step are never unmerged.

Definition 1

Let z_{ik} be vector of attribute values at site i assigned to cluster k whose size is n_k and mean is $\bar{z}_k$. Then the error sum of squares for cluster k is

$$E_k = \sum_{i=1}^{n_k} \|z_{ik} - \bar{z}_k\|^2 \qquad \square$$

Definition 2

Let E_k be the error sum of squares in Definition 1 and let g be the total number of clusters. The total within group error sum of squares is

$$E = \sum_{k=1}^{g} E_k \qquad \square$$

Table 1 and Figure 2 illustrate the method for a simple case involving a single attribute. At the beginning each measurement coincides with its centroid and both the error per cluster and consequently the total error are zero. In the first round of mergers, only the merging sites contribute to the total sum, so $E = E_k$.

For large numbers of clusters, instead of computing the new total sum of errors directly as the sum of all clusters in Definition 2, there are savings in the calculations by computing the new total sum of errors as the lowest sum from the previous round, plus the new error for the new merging clusters, minus the individual cluster errors for the merging clusters. Even larger savings arise by working directly with the square distances between centroids instead of the within square distances E_k (Anderberg, 1973, p. 142-145; Kendall, 1986, p. 37).

Table 1. Example illustrating the application of the Ward's method showing the error sum of squares and the total within group error sum of squares only for the underlined, new partitions plus the mean for the best mergers.

Number of clusters	Possible partitions of the measurements	E_k	E	$\bar{z}_k$
6	(0.1) (1) (3) (7) (8) (10)			
5	<u>(0.1, 1)</u> (3) (7) (8) (10)	0.405	0.405	0.55
	<u>(0.1, 3)</u> (1) (7) (8) (10)	4.205	4.205	
	<u>(0.1, 7)</u> (1) (3) (8) (10)	23.805	23.805	
	<u>(0.1, 8)</u> (1) (3) (7) (10)	31.205	31.205	
	<u>(0.1, 10)</u> (1) (3) (7) (8)	49.005	49.005	
	(0.1) <u>(1, 3)</u> (7) (8) (10)	2	2	
	(0.1) <u>(1, 7)</u> (3) (8) (10)	1	18	
	(0.1) <u>(1, 8)</u> (3) (7) (10)	24.5	24.5	
	(0.1) <u>(1, 10)</u> (3) (7) (8)	40.5	40.5	
	(0.1) (1) <u>(3, 7)</u> (8) (10)	8	8	
	(0.1) (1) <u>(3, 8)</u> (7) (10)	12.5	12.5	
	(0.1) (1) <u>(3, 10)</u> (7) (8)	24.5	24.5	
	(0.1) (1) (3) <u>(7, 8)</u> (10)	0.5	0.5	
	(0.1) (1) (3) <u>(7, 10)</u> (8)	4.5	4.5	
	(0.1) (1) (3) (7) <u>(8, 10)</u>	2	2	
4	<u>(0.1, 1, 3)</u> (7) (8) (10)	4.407	4.407	
	<u>(0.1, 1, 7)</u> (3) (8) (10)	28.14	28.14	
	<u>(0.1, 1, 8)</u> (3) (7) (10)	37.407	37.407	
	<u>(0.1, 1, 10)</u> (3) (7) (8)	59.94	59.94	
	(0.1, 1) <u>(3, 7)</u> (8) (10)	8	8.405	
	(0.1, 1) <u>(3, 8)</u> (7) (10)	12.5	12.905	
	(0.1, 1) <u>(3, 10)</u> (7) (8)	24.5	24.905	
	(0.1, 1) (3) <u>(7, 8)</u> (10)	0.5	0.905	7.5
	(0.1, 1) (3) <u>(7, 10)</u> (8)	4.5	4.905	
	(0.1, 1) (3) (7) <u>(8, 10)</u>	2	2.405	
3	<u>(0.1, 1, 3)</u> (7, 8) (10)	4.407	4.907	1.367
	<u>(0.1, 1, 7, 8)</u> (3) (10)	49.208	49.208	
	<u>(0.1, 1, 10)</u> (3) (7, 8)	59.94	60.44	
	(0.1, 1) <u>(3, 7, 8)</u> (10)	14	14.405	
	(0.1, 1) <u>(3, 10)</u> (7, 8)	24.5	24.905	
	(0.1, 1) (3) <u>(7, 8, 10)</u>	4.667	5.077	
2	<u>(0.1, 1, 3, 7, 8)</u> (10)	50.048	50.048	
	<u>(0.1, 1, 3 , 10)</u> (7, 8)	60.308	60.808	
	(0.1, 1, 3) <u>(7, 8, 10)</u>	4.667	9.074	8.333
1	<u>(0.1, 1, 3, 7, 8, 10)</u>	81.875	81.875	4.85

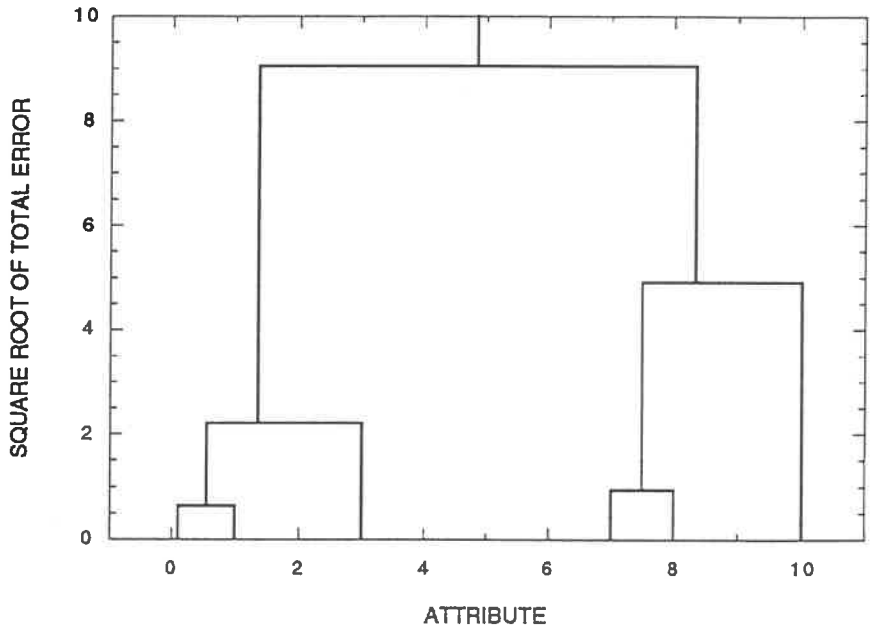


Figure 2. The Ward's method tree for the example in Table 1.

Algorithm 1

This is an algorithm to perform cluster analysis using the Ward's method.

1. Start with as many clusters as sites, each cluster consisting of exactly one site. At this stage the value of E in Definition 2 is zero.
2. Reduce the number of clusters by one by merging those two that minimize the increase of the total error E.
3. If the sites are in more than one cluster, go back to step 2.
4. Display the results in the form of an inverted tree showing at each stage which two clusters were merged and its corresponding total error E or total number of clusters g. ❑

Most statistical packages include implementations of several methods for cluster analysis, including Ward's (IMSL, 1987; SAS Institute, 1990). Among the properties of Ward's method one has:

(a) The method is biased toward finding clusters with the same number of sites in each cluster (SAS, 1990, p. 56).

(b) Because sites clustered at a previous clustering step are never unmerged, the calculations are relatively simple, but they often result in suboptimal partitions conditioned to the previous steps. For example, at the three cluster level the method produces the clusters (1, 2), (4, 6), (9, 10) for this univariate sampling of size 6. In the next stage the method produces the clusters (1, 2, 4, 6) and (9, 10) with total error $E = 15.25$, which is larger than the smallest total error

$E = 13.333$ associated with clusters (1, 2, 4) and (6, 9, 10), an unfeasible partition after selecting (1, 2), (4, 6), (9, 10) in the previous stage.

Despite its suboptimality, the method has been quite successful reproducing clusters in blind tests.

(c) Ward's method is unable to avoid calculations for the partitions involving small clusters when they are of no interest. For example for a sampling of size 1,000, if the interest is restricted to partitions with 5 to 20 clusters, it is possible to stop the calculations after completing the partition into 5 clusters, but it is not possible to avoid the more numerous calculations for partitioning the sampling from 999 to 21 clusters.

(d) The total sum of errors increases monotonically as the number of clusters decreases.

(e) The method is not robust with respect to survey errors and outliers (Mojena, 1988).

DISCRIMINANT ANALYSIS

In practice the answer to the question of how to group vectorial observations z is never straightforward. First there is the problem of deciding how many groups, which to a large extent one can settle in terms of external information or rules based on the total sum of errors that one can obtain running cluster analysis.

A second and more difficult problem is the group assignment for observations that are not like any of the typical observations in a group and at the same time poorly resemble more than one group. Cluster analysis assignment for such problematic observations is unstable and varies depending on the measure of similarity and the clustering method.

If the quality of group assignments is a concern, cluster analysis does not offer easy answers. That is the realm of discriminant analysis, which decides the assignment of any vector z based on allocation probabilities. Discriminant analysis, however, requires a training classification, which in regionalized classification is the one generally produced by cluster analysis.

The basis of discriminant analysis for assigning observations to one of a given number of groups is the minimization of the total misallocation cost. The procedure assumes that it is possible to partition the attribute space in as many mutually exclusive and exhaustive regions R_i as there are groups. A site is said to belong to group i when the vector z of attributes associated with the site falls in R_i.

Theorem 1

Let $\mathbf{Z}$ be a vectorial random function with probability density function $f(z)$ and let $z(x)$ be a realization of $\mathbf{Z}$ at site x, z for short. Let π_i be the proportional share of observations in the ith group whose probability density function is $f_i(z)$.

Then the probability $p_i(z)$ that the site characterized by z belongs to the ith group is

$$p_i(z) = \frac{\pi_i f_i(z)}{f(z)}$$

Proof

The proof directly follows from Bayes' Theorem. If one considers that π_i can be regarded as the a priori probability of sampling group i, then $p_i(z)$ is the posterior probability that z belongs to group i. ❑

The optimality in terms of the misallocation cost is assured by assigning z to the group with the highest probability (McLachlan, 1992, p. 7).

The kernel of discriminant analysis is the calculation of the probabilities in Theorem 1. As is always the case in statistics, there are non-parametric and parametric methods. By far, multivariate normal methods prevail among the latter and overall.

Definition 3

Let $z(x)$ be a realization of a vectorial random function at site x, z for short. Let $\bar{z}_i$ be the vectorial mean and Σ_i the covariance matrix of group i. Then the Mahalanobis distance from z to the centroid of group i is the squared weighted distance

$$\delta_i^2(z) = (z - \bar{z}_i)\Sigma_i^{-1}(z - \bar{z}_i) \qquad ❑$$

Theorem 2

Let $\boldsymbol{Z}$ be a normal vectorial random function with heteroscedastic normal group distributions. If π_i is the a priori probability of group i and $\delta_i^2(z)$ is the Mahalanobis distance in Definition 3, then the probability $p_i(z)$ that the site characterized by z belongs to the ith group is

$$p_i(z) = \frac{\pi_i |\Sigma_i|^{-1/2} e^{-\frac{\delta_i^2(z)}{2}}}{\sum_{i=1}^{g} \pi_i |\Sigma_i|^{-1/2} e^{-\frac{\delta_i^2(z)}{2}}}$$

Proof

From Theorem 1

$$p_i(z) = \frac{\pi_i f_i(z)}{f(z)}.$$

If the probability distribution for the ith group is normal,

$$f_i(z) = (2\pi)^{-p/2} |\Sigma_i|^{-1/2} e^{-\frac{\delta_i^2(z)}{2}}$$

where p is the dimension of the attribute space. Considering that in addition $f(z) = \sum_{i=1}^{g} \pi_i f_i(z)$ (McLachlan, 1992, p. 5), then

$$p_i(z) = \frac{(2\pi)^{-p/2} \pi_i |\Sigma_i|^{-1/2} e^{-\frac{\delta_i^2(z)}{2}}}{\sum_{i=1}^{g} (2\pi)^{-p/2} \pi_i |\Sigma_i|^{-1/2} e^{-\frac{\delta_i^2(z)}{2}}}.$$

Cancellation of the $(2\pi)^{-p/2}$ terms proves the theorem. ❑

Theorem 3

Let $\mathbf{Z}$ be a normal vectorial random function with homoscedastic normal group distributions. If π_i is the a priori probability of group i and $\delta_i^2(z)$ is the Mahalanobis distance in Definition 3 computed using the common covariance matrix $\boldsymbol{\Sigma}$, then the probability $p_i(z)$ that the site characterized by z belongs to the ith group is

$$p_i(z) = \frac{e^{\ln(\pi_i) - 0.5\delta_i^2(z)}}{\sum_{i=1}^{g} e^{\ln(\pi_i) - 0.5\delta_i^2(z)}}$$

Proof

From Theorem 2, considering that all covariance matrices are the same

$$p_i(z) = \frac{\pi_i |\Sigma|^{-1/2} e^{-\frac{\delta_i^2(z)}{2}}}{|\Sigma|^{-1/2} \sum_{i=1}^{g} \pi_i e^{-\frac{\delta_i^2(z)}{2}}}$$

Proof follows by the cancellation of the covariance matrix determinant term and consolidation of exponents after the trivial transformation $\pi_i = e^{\ln(\pi_i)}$. ❑

Use of the model in Theorem 3 results in hyperplanes for the boundaries of the group regions R_i, thus the term of linear discriminant analysis, while quadratic discriminant analysis deals with different covariance matrices and second order surfaces arising from the heteroscedastic model.

Whatever the model of discriminant analysis used for regionalized classification, the model requires a training set to determine the number of groups and for the assignment of realizations to the different groups in order to have some data for estimating the centroid and the covariance matrix of such groups.

Algorithm 2

This is a procedure for the calculation of group probabilities for vectorial measurements. The method employs a normal model of discriminant analysis.

1. Run Algorithm 1 using the whole sampling of the coregionalization.

2. Break the sampling into groups based on the total sum of errors, some external criteria, or both.

3. Use the vectorial measurements to estimate all group centroids, covariance matrices, and make a decision about the a priori group probabilities. The alternatives are:

   (a) If the sampling properly represents the true group sizes, the best estimate of the a priori probabilities are the relative proportions of the n_i sites per group

$$\pi_i = n_i / \sum_{i=1}^{g} n_i$$

   (b) If the proportions of sites per group are regarded as not being indicative of the a priori group probabilities and the user believes that he or she has some better external information to assess the probabilities, the user may employ such external assessment.

   (c) If the proportions of sites per group are regarded as not being indicative of the a priori group probabilities and the user lacks ways to assess them, they can be made equal to $1/g$, in which case they are ignored in the calculations.

4. Compare the group covariance matrices and decide whether they are sufficiently similar to assume homoscedasticity.

5. Calculate the Mahalanobis distance in Definition 3 for measurement z. Use the average of all group covariances if the assumption is that the covariance is homoscedastic.

6. Compute for measurement z the probability $p_i(z)$ of belonging to each of the groups

$$p_i(z) = \frac{e^{-0.5D_i^2(z)}}{\sum_{i=1}^{g} e^{-0.5D_i^2(z)}}$$

where, if

$D_i^2(z) = \delta_i^2(z)$, the discriminant analysis is linear with unknown or equal a priori group probabilities;

$D_i^2(z) = \delta_i^2(z) - 2\ln(\pi_i)$, the discriminant analysis is linear with known and different a priori group probabilities;

$D_i^2(z) = \ln|S_i| + \delta_i^2(z)$, the discriminant analysis is quadratic with unknown or equal a priori group probabilities;

$D_i^2(z) = \ln|S_i| + \delta_i^2(z) - 2\ln(\pi_i)$, the discriminant analysis is quadratic with known and different a priori group probabilities.

The choice of $D_i^2(z)$ must be consistent with the decision of homoscedasticity made in step 4. The term $\delta_i^2(z)$ is the Mahalanobis distance in Definition 3; and S_i is the estimate of the covariance matrix for group i;

7. Go back to step 7 until assigning group probabilities for all sites in the training set. ❑

Major statistical computer packages such as IMSL(1987) and SAS (1990) have implementations of Algorithm 2 among several other procedures.

Statisticians concur that quadratic discriminant analysis indeed provides superior results if the group covariances are considerably different and the group sizes are large. Quadratic discriminant analysis, however, is more sensitive to deviations from multinormality and assignment errors in the training set (Lachenbruch, 1982).

ALLOCATION BY EXTENSION

The use that regionalized classification makes of discriminant analysis is to a certain degree analogous to the use of crossvalidation in kriging and different from the classical use of discriminant analysis. Classical use of discriminant analysis employs the training set for calibration and then proceeds to classify vectorial measurements without assigments.

In regionalized classification the interest is in the calculation of the probabilities $p_i(z)$ for the same realizations in the training set already classified by the prior cluster analysis. Once one has all the probabilities $p_i(z)$, reallocation of the sites to the group with the highest probability is a trivial endeavor. This reallocation offers an opportunity to check results and compare methods in case the user wants to consider more than one type of cluster analysis or discriminant analysis. Results from cluster and discriminant analysis should be comparable only with minor variations that one should employ to select methods and parameters yielding the most consistent results.

The final step in regionalized classification is the mapping of groups, which one can accomplish by arbitrarily assigning colors or black and white patterns to the groups. As arbitrary as the color or pattern selection may be, it always helps to select a combination of alternatives that maximizes contrast to the eye, which customarily requires some experimentation by trial and error.

Most mapping procedures require a collection of values regularly spaced at short intervals, which is rarely the case of training sets. In addition, interpolation presumes that the variable is continuous. In those circumstances Harff and Davis (1990) recommend mapping the group probabilities $p_i(z)$ by treating them as regionalized variables and then use the allocation rule to produce the discontinuous group map. Remember that z is a shorthand for $z(x)$, where x is the location of the site. Then the probabilities are actually regionalized variables $p(z(x))$ or $p(x)$ that depend on location. One does group

allocation node by node performing a grid-to-grid operation in which one assigns each node to the group with the highest probability.

Considering that each site must be fully sampled to allow for the cluster and discriminant analysis, there is no advantage on using cokriging for the estimations. Although the probabilities sum to a constant, use of alr transformation is not feasible due to the numerous zeros both in the numerator resulting in a null argument for the logarithm, or in the denominator producing unacceptable ratios.

Ordinary or universal kriging, the default geostatistical options of choice, suffer from the inability to restrict the estimates to an interval—0-1 in this instance—let alone to force the probabilities to sum to one. Estimates outside the 0-1 interval, however, are rare and never far away from the interval. A common solution to force a vector in a coregionalization to add to one is to rescale the values. Considering that such correction does not change the ranking of the membership probabilities, the allocation is insensitive to the rescaling.

Alternative strategies involving the interpolation of the coregionalization itself or of the generalized distances instead of the membership probabilities are deceptive choices. Although the probabilities end up honoring all constraints, the use of estimates instead of true values results in unaccounted propagation of errors.

Allocation in regionalized classification remains open to improvements.

Algorithm 3

This is a procedure for the regionalized classification of fully sampled coregionalizations involving p attributes.

1. Assign the sites to one and only one of g groups either by using Algorithm 1 or any other cluster analysis procedure, or on the basis of external information
2. For each site calculate the group probabilities either by Algorithm 2 or any other discriminant analysis procedure deemed appropriate.
3. If the mapping procedure does not require a regular grid of values, go to step 4. Otherwise, use some form of kriging to produce grids of estimated values for every group probability.
4. For every site or node, assign the site or node to the group with the largest probability.
5. Prepare a map showing the group assignment. This is the regionalized classification of the area under study, conditional to the values and attributes in the sampling.
6. Prepare a map of the highest probability per node. This is the probability that the site in the regionalized classification indeed belongs to the group in the regionalized classification.
7. Prepare a censored map eliminating nodes or sites likely to be misclassified, which are more likely to occur when the highest probability is only marginally larger than the second highest probability. ❑

THE DAKOTA AQUIFER CASE STUDY

The Dakota aquifer is a complex unit schematically shown in Figure 3. Spatial fluctuations as basic as the variation in the amount of shale are important for the understanding of the geology and the modeling of fluid flow. Perennial lack of information at close spacing is responsible for the poor understanding of heterogeneity across the aquifer.

The sampling in the appendix is a first attempt to overcome lack of data at least at the county scale. Data in the appendix is a by-product of well logging by the petroleum industry exploiting fields beneath the aquifer. As shown in Figure 4, the study area is a 3 by 3 township square in south central Hodgeman county.

Experimentation showed us that the Dakota Formation is too thick a unit for the purpose of characterizing its heterogeneity. In addition the uppermost part of the formation is missing due to fluvial erosion in the eastern side of the study area. Following the practice in neighboring Colorado, we divided the formation into the upper "D" and the lower "J" sandstones, discarding the partly missing "D" and subdividing the thicker "J" sandstone into the basal 120 ft and the remainder of the deposits above. When present, the "D" sandstone in the area ranges from 20-90 ft. 331 gamma ray logs were digitized and entered into a data base that was then accessed by a computer program.

Combined laboratory and field work has demonstrated that groundwater flow occurs at those levels where the natural gamma ray radiation in cased wells is below 60 API units (P. A. MacFarlane, 1995, personal communication). Hence we considered as sandstone any sediment below that radiation threshold in cased wells and below a 55 API threshold in uncased wells. Further experimentation to find the most revealing attributes found that cumulative sandstone thickness was the best choice. The computer program automatically accounted the cumulative amount of sandstone per unit. The appendix includes only those 215 wells with all four units fully logged—a requirement of regionalized classification—plus additional 16 wells not included in the well log data base for which we measured the amount of cumulative sandstone manually. Figure 5 contains maps for the cumulative sandstone thicknesses in all four units using all information per unit, which tends to be more abundant with depth due to the fact that logging of surface units tends to be skipped because of their irrelevance for the oil industry.

Well clustering by the Ward's method

For any real multivariate sampling it is not possible to prepare as detailed a tree as the one in Figure 2. Figure 6 is a possible alternative rendition restricting the attention to the final stages of the clustering, which customarily are the ones with the partitions of interest. A bar diagram with the mean value of the attributes per cluster is one way to reduce the dimensionality of the data. In Figure 6, the vertical axis goes from 0-150 ft and the thickness sequence from left to right is upper "J", lower "J", Kiowa, and Cheyenne cumulative sandstone thickness. The numerical value denotes the number of wells in the cluster.

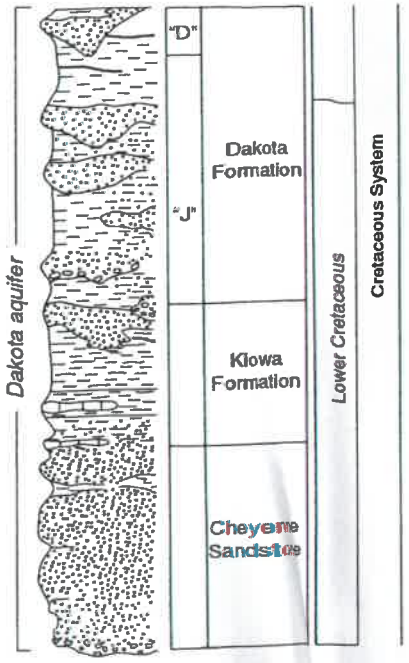


Figure 3. Generalized stratigraphic column for the Dakota aquifer in Kansas (After MacFarlane and Sawin, 1995).

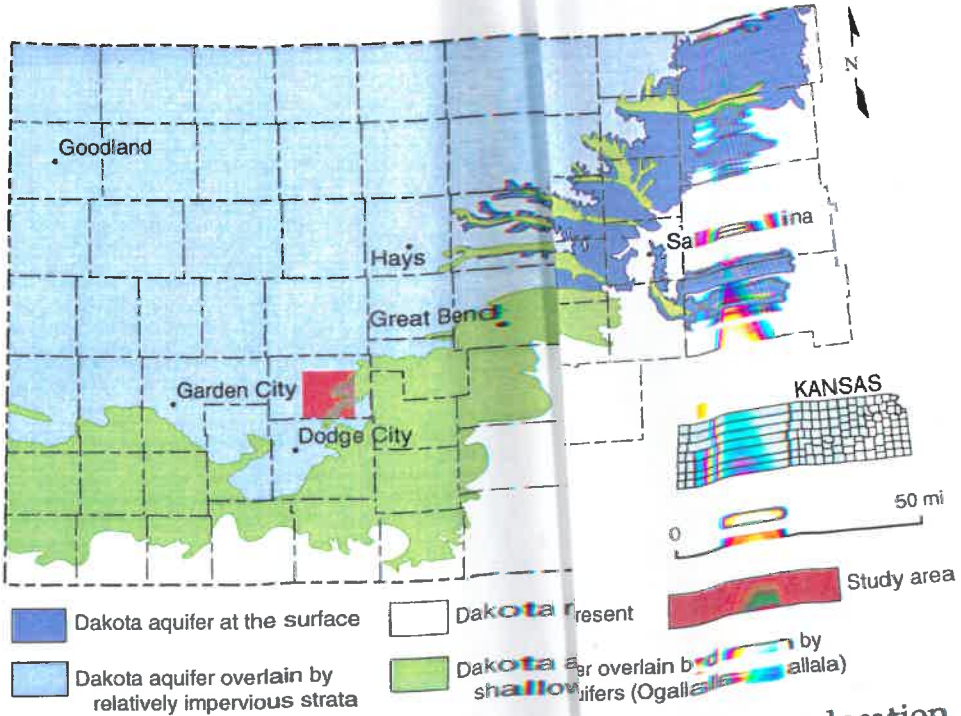
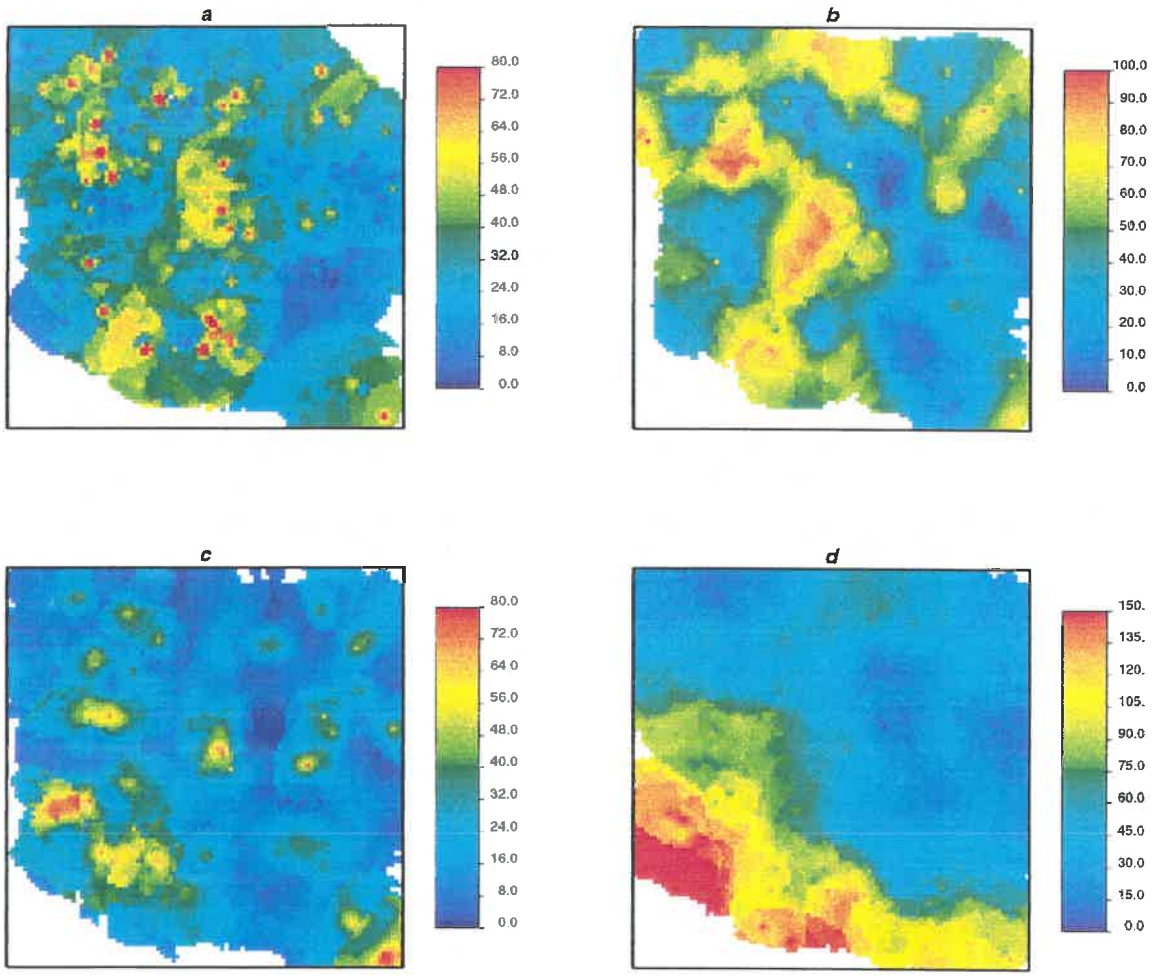


Figure 4. The Dakota aquifer in Kansas showing location of the study area (After MacFarlane and Sawin, 1995)



sandstone

Figure 5. Cumulative thickness maps for units comprising the Dakota aquifer, southcentral Hodgeman County, Kansas: (a) upper "J", sample size 231; (b) lower "J", sample size 294; (c) Kiowa, sample size 324; (d) Cheyenne, sample size 344.

Naming clusters for easy reference in the text is another challenge in the display of results. Here we decided to take advantage of the uniqueness of cluster size and label each cluster by its size.

Remember that Ward's method progresses from largest to smallest number of clusters. Traveling the tree in inverse direction from its generation one can observe that:

i. The topmost bar diagram provides a pictorial representation of the mean value per attribute for the whole sampling whose size is 231 in this case. The last merger is between a group of 49 wells with the rest of the 182 wells. The smaller cluster has the wells with the most sandstone, especially for the Kiowa.

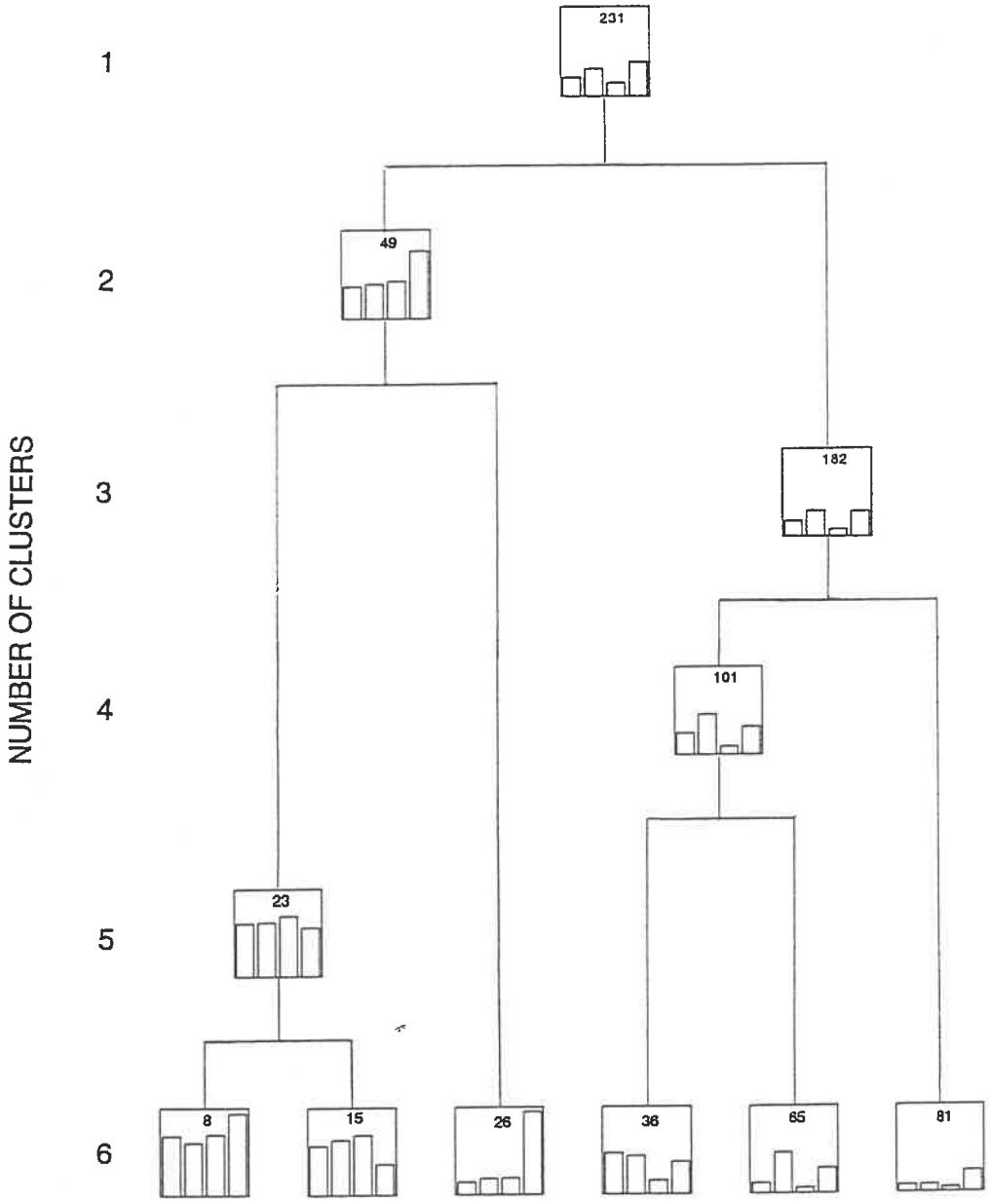


Figure 6. Final portion of the Ward's method tree for the training set of the case study. The horizontal axis does not have a scale and the boxes are bar diagrams for the coordinates of the centroids in a vertical scale 0-150 ft.

ii. Cluster 49 results from the merging of clusters 23 and 26 that mostly share large amounts of sandstone for the Cheyenne.

iii. At the three-cluster level Ward's method merges cluster 81 and 101 that basically share all average values except for the second attribute from the left—cumulative sandstone thickness for the lower "J".

iv. The previous four-cluster stage merges clusters 36 and 65, which differ only in the amount of upper "J" sandstone.

v. The five-cluster stage merges clusters 8 and 15 whose main discrepancy is the amount of cumulative sandstone in the Cheyenne.

vi. So far all breaking clusters relate to interesting characteristics on which to base the discriminant analysis in the next step. The six-cluster stage is no exception but the partition contains two clusters that are too small for a regionalized classification, thus bringing to a halt the interest on continuing analyzing the declustering of the tree.

The training set for discriminant analysis should be the partition into five clusters that have the sizes and means given in Table 2.

Table 2. Size, proportion, and means for the partition into five clusters

Group size	Proportion	Attribute	Mean
23	0.10	Upper "J" cum. ss	90.4
		Lower "J" cum. ss	92.8
		Kiowa cum. ss	102.9
		Cheyenne cum. ss	83.4
26	0.11	Upper "J" cum. ss	22.3
		Lower "J" cum. ss	28.1
		Kiowa cum. ss	29.0
		Cheyenne cum. ss	141.6
36	0.16	Upper "J" cum. ss	70.2
		Lower "J" cum. ss	65.0
		Kiowa cum. ss	23.6
		Cheyenne cum. ss	55.0
65	0.28	Upper "J" cum. ss	17.6
		Lower "J" cum. ss	69.6
		Kiowa cum. ss	9.4
		Cheyenne cum. ss	43.0
81	0.35	Upper "J" cum. ss	12.0
		Lower "J" cum. ss	12.3
		Kiowa cum. ss	7.5
		Cheyenne cum. ss	35.6

Discriminant analysis

According to the testing performed by the SAS (1990) program used in the processing, the within group covariance matrices in Table 3 are sufficiently different to make the quadratic discriminant analysis the model of choice.

Table 4 shows a 97% agreement between Ward's method clusters and the groups based on the posterior probabilities of membership. The group sizes are in the bottom line. Although they are slightly different than the corresponding cluster sizes, we will keep the cluster labels to avoid cluttering the exposition.

Table 3. Within-group covariance matrices for cumulative sandstone thickness.

Group		UJ	LJ	K	Ch
23	UJ	1532.1			
	LJ	344.6	615.9		
	K	-224.3	186.3	692.7	
	Ch	606.2	146.5	59.6	2312.4
26	UJ	317.1			
	LJ	53.2	490.8		
	K	127.3	140.5	965.0	
	Ch	-204.6	-17.9	-444.3	1543.1
36	UJ	436.8			
	LJ	317.6	1336.4		
	K	-73.1	214.0	330.7	
	Ch	-91.2	-176.3	21.3	511.1
65	UJ	254.9			
	LJ	57.3	329.8		
	K	-30.2	42.4	96.7	
	Ch	44.1	115.4	79.6	765.4
81	UJ	148.5			
	LJ	29.6	155.8		
	K	5.7	45.5	138.8	
	Ch	-39.7	29.8	97.1	616.5

Table 4. Resubstitution summary

	Group				
Cluster	23	26	36	65	81
23	23	0	0	0	0
26	0	25	0	1	0
36	0	0	33	1	2
65	0	0	0	65	0
81	0	2	0	1	78
Total	23	27	33	68	80

Considering that in this case the relative cluster sizes in the sampling are a good approximation to the true relative sizes, one should use the relative proportions in Table 2 as estimates for the a priori group probabilities.

Mapping of group membership probabilities.

Figure 7 shows the continuous variation for all membership probabilities across the study area. Table 5 contains the semivariogram models and Figure 8 displays both the models and the experimental semivariograms. To minimize extrapolation outside the convex hull of the sampling, estimates exceeding certain kriging variance threshold were ignored. From top to bottom and from left to right such maximum variance values for the maps in Figure 7 are 0.13, 0.055, 0.16, 0.235, 0.235.

Regionalized classification maps

Figure 9 displays the results of the grid allocations node by node. Remember that to minimize the misclassification error discriminant analysis allocates any vector of values to the group with the maximum group membership probability.

The group membership probabilities of the allocations are in Figure 10. They are the highest group membership probability per node.

Considering that the allocation is done for the most part based on estimated probabilities based on probabilities calculated by Algorithm 2 at the wells, membership probabilities away from wells may be in error. Hence close membership probabilities have potential for misclassification. The map in Figure 11 repeats the one in Figure 10 by eliminating those nodes whose difference between the highest and second highest probabilities is less than 0.1 times the sum of their kriging variances.

Table 5. Semivariogram models for group membership probabilities

Group	Type	Nugget	Sill	Range ft
23	Spherical	0.002	0.103	3137
26	Spherical	0.019	0.054	20251
36	Spherical	0.065	0.127	7140
65	Spherical	0.130	0.201	13050
81	Spherical	0.116	0.200	14104

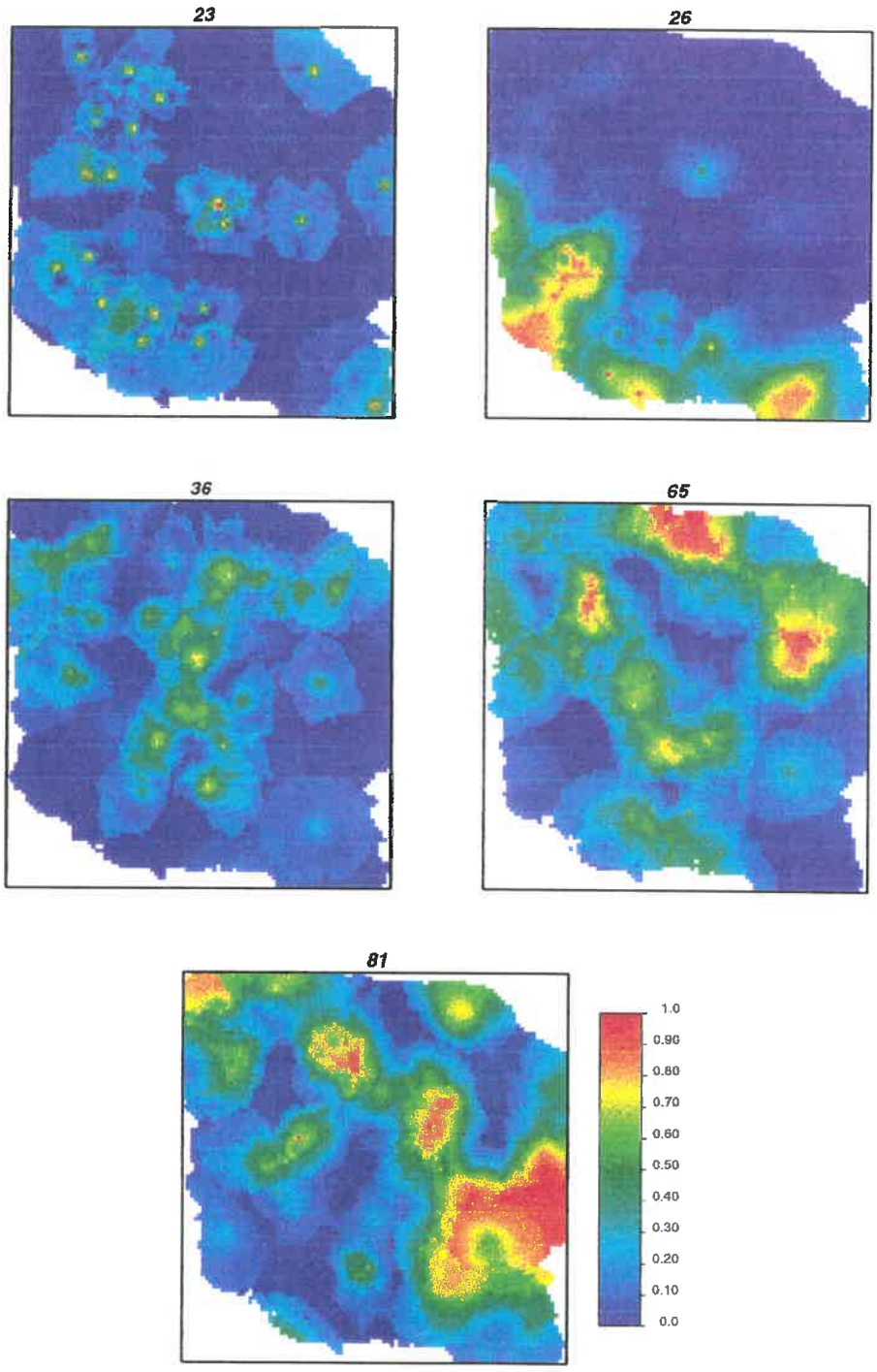


Figure 7. Membership probability maps for all five groups in the regionalized classification of the Dakota aquifer, southcentral Hodgeman County, Kansas.

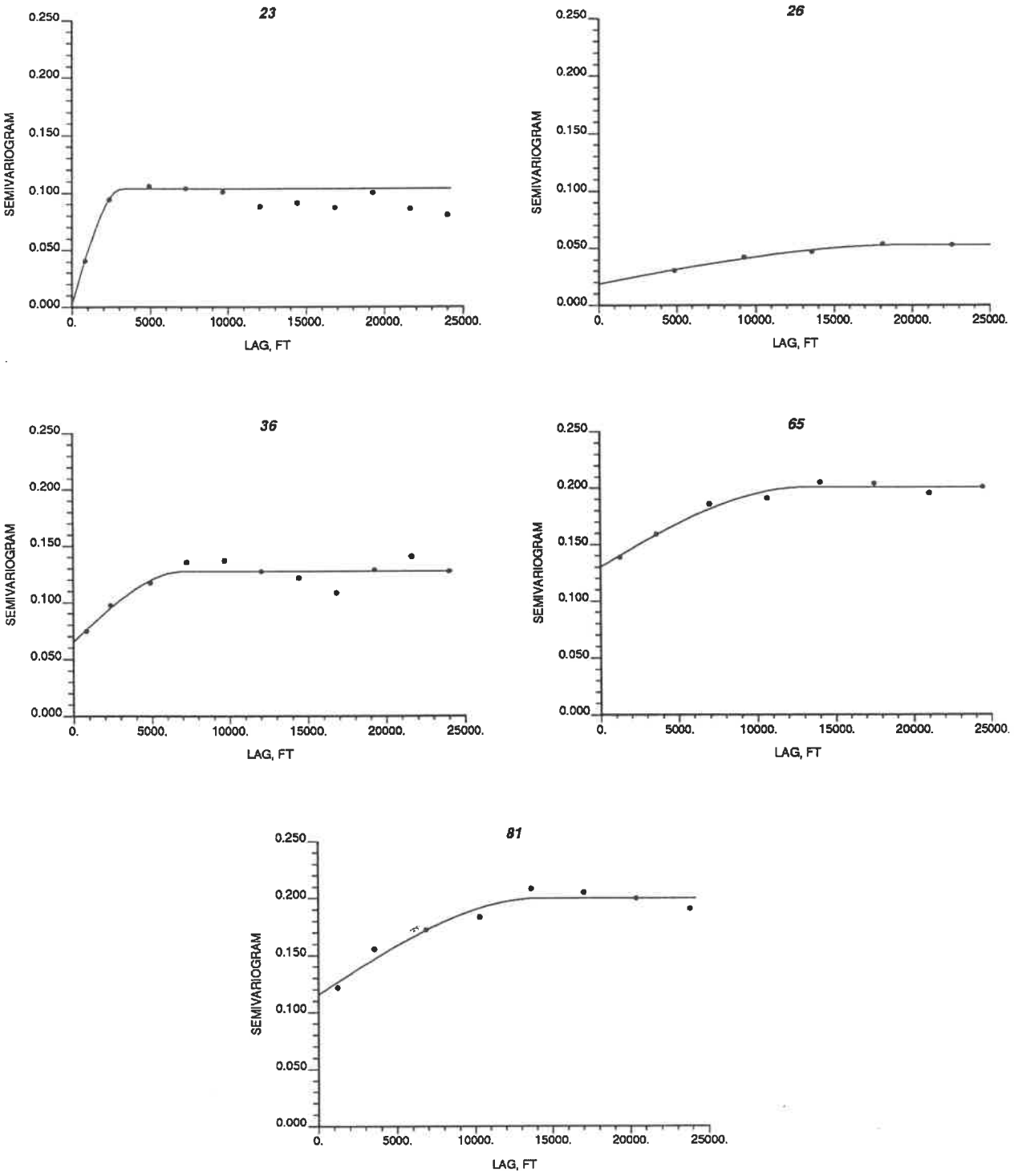


Figure 8. Probability semivariograms for all five groups in the regionalized classification of the Dakota aquifer, southcentral Hodgeman County, Kansas.

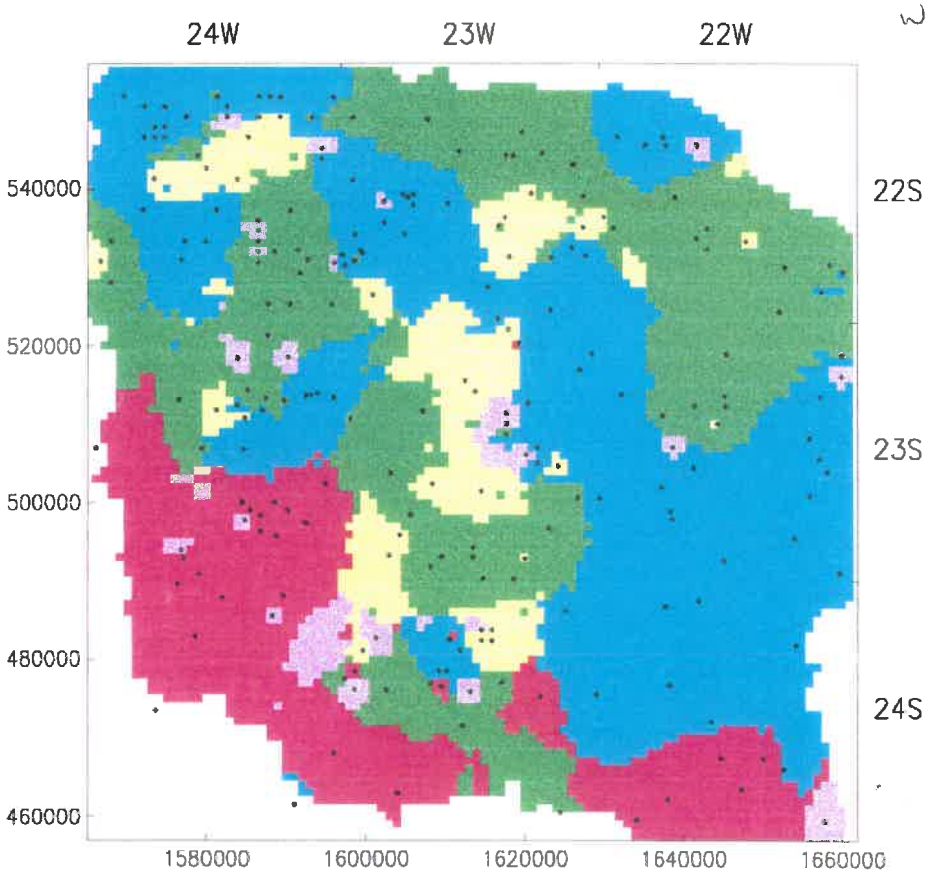


Figure 9. Regionalized classification of the Dakota aquifer, southcentral Hodgeman County, Kansas. Lavender denotes allocation to group 23, red to 26, yellow to 36, green to 65, and blue to 81.

CONCLUSIONS

Evaluation of the results of the regionalized classification requires borrowing from the basic geology of the Dakota aquifer and referring to either Figure 6 or Table 2 to remember the predominant characteristics of each group.

In the study area the aquifer is really two aquifers, the Cheyenne sandstone and the Dakota formation, which are separated by the Kiowa shale. The Cheyenne is saturated with salty water coming from the dissolution of underlying Permian salt beds while the Dakota formation is the unit of economic interest yielding fresh water. Throughout the area it is then important that the Cheyenne and the Dakota not be connected. At sites in group 23, a sandy Kiowa may result in an unavoidable leak in the sealing and one should be cautious with drilling and completions at sites in group 26 where the Cheyenne is thick.

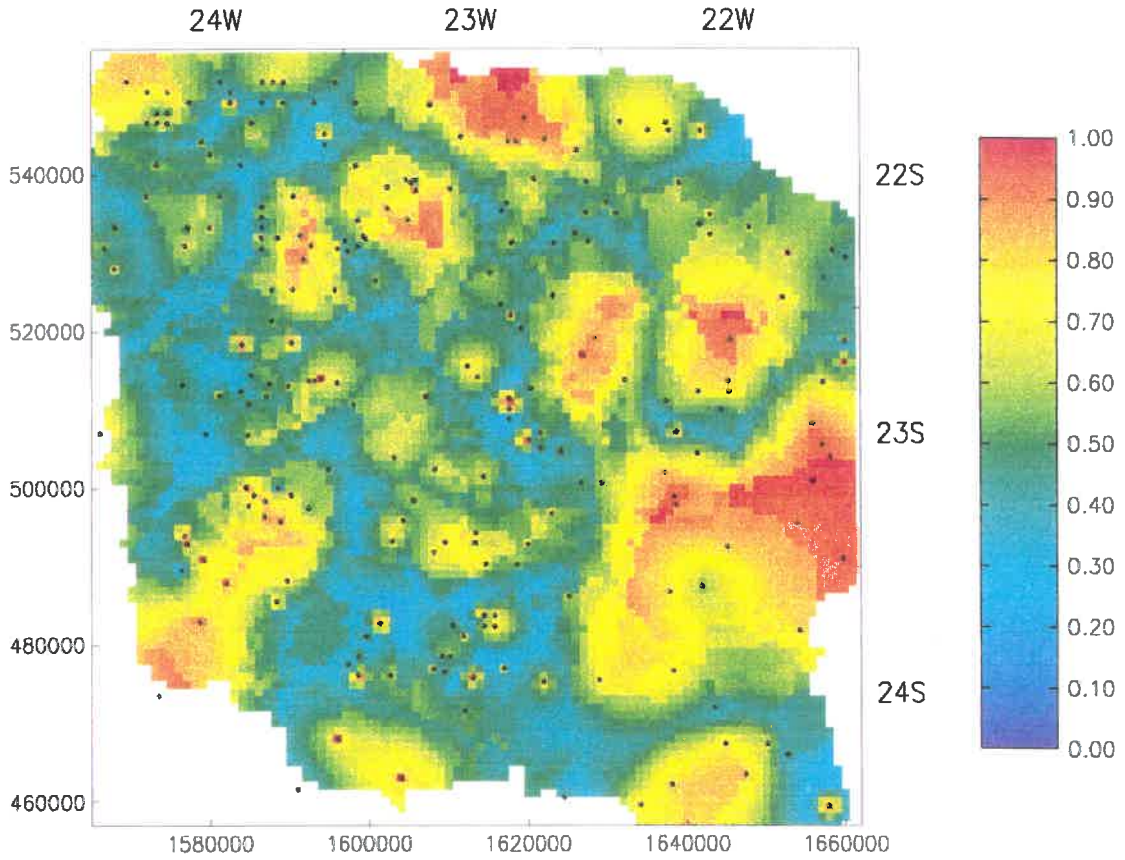


Figure 10. Group membership probabilities in the regionalized classification of the Dakota aquifer, southcentral Hodgeman County, Kansas.

Some of the group 23 anomalies may not be real, which would reduce the danger of leaks. There are grounds to believe that the gamma ray log scale reported in the logs may be incorrect for at least a few of those wells. Under-reported natural radiation results in units abnormally and incorrectly low in shale. The errors should not have a bearing upon the results of the classification of those wells outside group 23 and open another potential application of regionalized classification: detection of errors.

The third group to avoid is the most predominant group 81, where for some intriguing geological reason there has never been much deposition of sand. Thus the reasons to avoid drilling group 81 are economic rather than environmental.

Finally groups 36 and 65 should be primary targets for groundwater supply in the area. If one believes those geologists who postulate meandering river flood plains deposited the Dakota Formation, one may suggest that group 65 developed along a drainage system flowing to the northwest. The exceptionally favorable areas associated with group 36 tend to occur in the flanks of group 65.

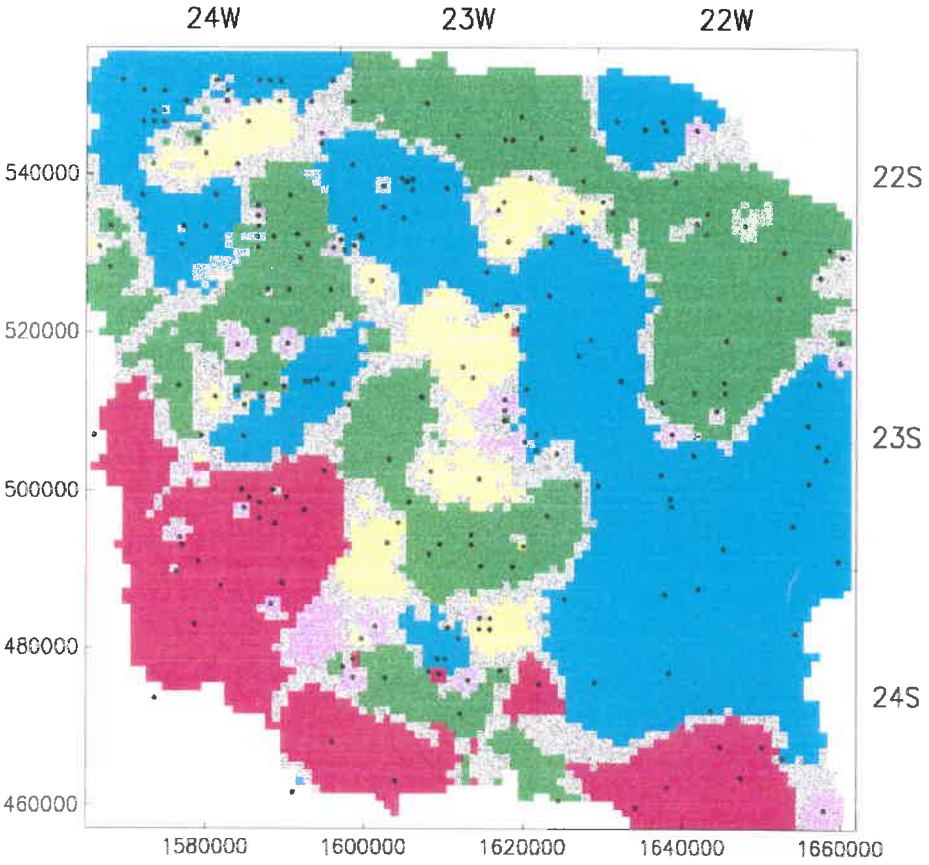


Figure 11. Regionalized classification of the Dakota aquifer, southcentral Hodgeman County, Kansas, eliminating maximum uncertainty allocations shown as gray areas. Lavender denotes allocation to group 23, red to 26, yellow to 36, green to 65, and blue to 81.

REFERENCES

Anderberg, M. R., 1973, Cluster Analysis for Applications: Academic Press, New York, 359 p.

Harff, J. and J. C. Davis, 1990, Regionalization in geology by multivariate classification: Mathematical Geology, vol. 22, no. 5, p. 573-588.

Kendall, M., 1980, Multivariate Analysis: Charles Griffin & Company Ltd., High Wycombe, England, second edition, 210 p.

Kogan, R. I., 1986, Interval'nye ocenki v Geologicheskich Issledovanijach: Nedra Press, Moscow, 335 p.

Lachenbruch, P. A., 1982, Discriminant analysis: in Encyclopedia of Statistical Sciences, volume 2, edited by S. Kotz and N. L. Johnson, John Wiley & Sons, New York, p. 389-397.

MacFarlane, P. A. and R. S. Sawin, 1995, A user's guide to well-spacing requirements for the Dakota aquifer in Kansas: Kansas Geological Survey, Public Information Circular 1, 4 p.

McLachlan, 1992, Discriminant Analysis and Statistical Pattern Recognition: John Wiley & Sons, Inc., New York, 526 p.

Mojena, R., 1988, Ward's clustering algorithm: in Encyclopedia of Statistical Sciences, volume 9, edited by S. Kotz and N. L. Johnson, John Wiley & Sons, New York, p. 529-532.

Rodionov, D. A., 1981, Statisticheskie Rezhenija v Geologii: Nedra Press, Moscow, 231 p.

SAS Institute, 1990, SAS/STAT User's Guide, Version 6, volume 1: SAS Institute Inc., Cary, North Carolina, 943 p.

Voronin, J. A., 1967, Geologija i Matematika: Nauka Press, Novosibirsk, 253 p.

APPENDIX

Table A.1. Dakota aquifer sampling used in the regionalized classification. The labels are taken from the original Stratamodel data base.

			Cumulative sandstone thickness			
ID	Easting ft	Northing ft	Upper "J" ft	Lower "J" ft	Kiowa ft	Chey-enne ft
5	1631797	546606	0.0	27.0	0.0	54.4
9	1635275	545570	13.0	9.0	2.0	58.3
12	1641860	545503	95.0	112.0	83.3	68.1
31	1630041	536362	33.0	96.0	56.3	49.1
32	1631291	535016	28.5	80.0	20.5	14.5
38	1643084	532306	17.9	51.0	8.9	30.5
40	1643137	534940	2.0	42.0	2.0	23.6
42	1648014	533313	73.0	70.0	32.6	60.1
48	1658548	530399	53.7	75.0	8.9	47.2
49	1660208	529439	6.0	34.0	8.9	52.1
50	1657520	526895	22.0	6.0	0.0	19.7
57	1652223	524384	2.1	58.0	25.8	21.6
59	1619712	547175	0.0	69.5	0.5	38.5
60	1607855	548835	16.0	94.0	0.0	33.0
61	1598543	549051	0.5	96.0	0.0	21.0
64	1611762	544744	38.3	84.0	1.5	70.5
68	1618670	544228	8.5	90.0	39.0	69.5

69	1617687	544248	1.5	63.0	0.0	33.0
70	1622289	544483	26.5	89.0	9.5	68.0
72	1626236	543077	46.5	73.0	0.0	50.0
75	1627484	539064	19.0	74.0	23.5	31.0
76	1620905	539390	95.5	55.5	11.0	37.0
79	1617557	536348	91.0	91.5	47.5	43.0
80	1610319	538152	0.5	0.0	0.0	50.0
81	1604730	539296	0.0	0.5	0.0	43.0
82	1606034	539265	56.5	11.5	16.0	37.0
83	1598477	541111	16.0	16.5	12.5	52.5
84	1602404	538352	120.0	102.5	76.0	
85	1602364	535691	31.0	19.0	23.0	49.5
87	1598759	534150	1.0	14.5	11.0	51.0
88	1599379	532145	25.5	25.5	8.0	40.5
89	1604990	534296	3.0	11.0	0.0	32.0
91	1616880	535363	5.5	45.5	10.5	44.5
92	1618118	531339	62.5	51.0	36.5	55.0
93	1623388	531203	3.0	69.0	0.0	4.0
94	1626054	532477	5.0	0.0	5.0	7.5
95	1627411	535098	48.0	74.5	53.0	44.0
97	1627694	531443	18.5	7.0	0.0	0.0
101	1615428	527431	36.0	0.0	0.0	32.5
104	1598727	530844	38.5	58.0	0.0	35.5
105	1600941	526476	73.0	92.0	27.5	80.5
106	1616657	523435	34.5	2.5	3.0	24.5
107	1617968	522083	105.0	78.5	21.0	23.0
108	1619267	520414	3.5	6.5	58.5	53.0
109	1623278	524586	40.5	14.0	18.5	37.5
112	1593222	549092	24.0	38.5	0.0	30.5
113	1595911	551702	20.5	31.0	0.0	8.0
114	1589279	551731	40.0	22.0	22.0	23.0
115	1587945	551746	2.5	4.5	5.5	9.0
116	1589238	549114	117.0	116.0	20.0	32.0
117	1586593	549133	0.0	41.5	0.0	0.0
118	1582610	549150	63.0	68.0	95.0	30.5
119	1581312	551771	15.0	83.0	3.5	0.0
120	1577424	549170	16.5	37.5	30.0	33.0
121	1573371	547856	1.0	54.0	9.5	22.5
122	1574698	550486	38.5	9.0	0.0	15.5
123	1572083	550477	19.0	22.0	22.5	34.5
124	1574674	547858	19.0	57.0	12.5	18.5
127	1573367	546534	25.0	20.5	19.0	32.0
128	1574670	546535	23.5	13.5	19.5	21.5
129	1572038	546533	15.5	35.0	0.0	21.5
130	1579926	542551	94.0	109.0	10.0	51.0
131	1578934	544212	40.0	66.0	0.0	11.0

132	1585232	546514	80.0	108.5	0.0	0.0
133	1594535	545127	17.5	73.5	90.0	26.5
134	1594533	543799	0.0	44.0	46.5	35.0
135	1590518	537219	13.0	67.5	0.0	13.5
136	1583908	541206	37.5	31.0	0.0	67.5
138	1581226	537274	8.5	5.0	0.0	47.0
139	1573283	541245	81.5	111.0	10.0	21.5
140	1571907	537271	9.0	4.5	0.0	47.0
141	1567911	533340	31.0	75.5	0.0	27.5
142	1579869	533288	17.0	9.0	6.5	49.5
143	1577213	533286	64.5	4.0	48.5	53.5
144	1586525	535925	40.5	66.5	19.0	41.5
145	1586523	534597	62.0	69.0	88.5	37.5
146	1586502	533263	68.0	4.5	9.5	52.0
147	1586499	531941	154.0	117.5	116.5	46.5
148	1591460	532243	22.0	71.5	12.0	29.0
149	1588483	531932	7.5	108.5	34.0	31.0
150	1597100	531552	42.5	65.0	1.5	33.5
151	1592775	530917	0.0	104.5	17.0	35.0
152	1596115	530561	46.0	86.5	96.5	40.5
153	1597431	530224	3.0	3.0	0.0	35.0
154	1591802	529276	16.0	84.0	0.5	30.5
155	1586496	530624	38.0	53.0	10.5	13.5
156	1576865	530969	0.0	1.0	0.0	24.0
158	1566564	530705	52.5	108.5	50.0	59.0
159	1567861	528039	19.5	116.0	6.0	65.0
162	1587693	521375	3.0	89.5	41.5	74.5
163	1590455	525323	41.5	94.5	1.5	27.0
164	1587754	525342	111.5	99.0	20.5	71.5
165	1595746	525270	0.5	66.0	12.5	52.0
166	1660074	518849	0.0	37.0	0.0	1.0
167	1660023	516188	53.0	92.5	104.5	22.5
170	1645547	518976	0.0	51.0	0.0	0.0
172	1632313	513836	0.0	0.0	0.0	0.0
173	1637517	511137	46.0	55.5	1.0	31.0
174	1644409	510045	61.2	31.0	47.0	48.0
175	1645310	513673	10.0	70.5	17.5	21.5
176	1645442	512348	0.5	78.0	25.5	35.5
180	1657332	513561	1.5	5.0	1.0	2.0
181	1657231	505627	18.5	0.0	0.0	9.5
182	1655933	508283	0.0	0.0	0.0	0.0
183	1658211	503959	23.0	0.0	0.0	8.5
186	1641382	504456	0.0	0.0	0.0	8.5
187	1638777	507136	67.5	38.5	125.0	26.0
188	1629429	500645	11.0	6.0	0.0	32.5
190	1637376	501970	17.5	0.0	1.5	15.0

194	1645180	492553	0.0	0.0	0.0	20.0
196	1659722	491076	40.5	12.0	0.0	14.5
200	1612504	515610	59.5	82.5	30.5	42.5
201	1598092	510693	9.5	43.5	6.0	66.0
202	1607242	511741	10.0	80.0	0.0	53.0
203	1617728	510145	97.0	113.5	117.5	60.0
204	1613780	514247	75.0	84.5	12.5	59.0
206	1620423	512752	28.0	0.0	6.5	41.0
209	1624242	504707	85.5	108.0	43.0	42.0
210	1620180	506103	79.0	94.5	95.5	53.5
211	1621621	507067	24.0	15.0	0.0	17.5
212	1621629	505088	8.5	25.5	0.5	28.0
213	1617709	508811	36.0	49.5	4.0	39.5
214	1603136	503894	40.5	73.5	7.5	71.0
215	1608410	502405	79.0	112.0	22.5	72.5
217	1626793	500687	24.5	94.0	0.0	15.5
219	1623072	496783	13.0	61.5	5.5	36.0
220	1613493	494335	12.0	75.5	16.0	62.0
221	1605694	498516	15.5	68.5	18.5	61.0
224	1602927	493317	50.5	88.0	37.0	70.0
225	1608186	491845	24.5	58.0	4.5	68.5
226	1609537	493126	17.5	71.0	20.5	71.0
227	1614707	490372	26.5	53.5	1.5	39.0
228	1618663	490272	20.0	61.0	16.0	30.5
229	1620032	492875	72.5	53.0	23.5	27.0
230	1590299	518562	120.5	116.5	107.5	111.0
231	1583971	518421	92.0	102.5	87.0	82.5
233	1581198	511851	56.3	38.5	29.0	87.5
234	1576550	513236	2.0	61.3	4.0	110.5
235	1584833	510794	58.0	61.0	25.0	98.5
236	1585197	514429	0.0	49.0	14.0	79.5
237	1583875	512471	4.0	11.0	8.0	71.5
239	1587483	513410	3.0	68.0	8.5	92.5
240	1589811	513091	3.5	51.5	16.0	74.5
241	1587027	511761	11.5	0.0	9.0	87.0
242	1593182	513751	12.5	22.5	18.5	91.0
243	1592497	513761	4.5	12.0	31.5	89.5
244	1584688	506796	0.0	3.0	0.0	61.5
245	1579420	506900	40.5	37.0	0.0	49.0
246	1584514	500119	34.0	27.0	49.5	126.5
247	1588487	500097	11.5	0.0	0.0	44.0
248	1594927	502433	39.0	60.5	86.5	101.0
249	1592456	497486	29.0	0.0	3.0	93.0
250	1590160	499115	42.5	19.5	16.5	115.5
251	1588760	495803	11.0	32.5	39.5	124.5
252	1586747	496434	9.0	7.0	31.5	130.5

253	1586795	498411	53.0	26.0	4.5	133.5
254	1585465	499099	3.0	25.0	42.5	118.5
257	1579042	490924	11.5	30.0	5.0	126.5
258	1576774	493989	34.0	80.5	135.5	177.5
259	1577070	492973	0.5	44.0	35.0	164.5
260	1576349	489633	0.0	31.5	0.0	41.0
262	1637835	486794	2.5	10.5	43.0	39.5
272	1638333	476736	26.5	0.0	0.0	0.0
273	1628959	475530	0.0	0.0	0.0	15.5
275	1644813	467420	46.0	42.5	21.0	124.0
277	1650124	467367	62.0	16.0	117.5	98.5
279	1652711	466027	7.5	1.0	23.0	92.0
280	1647401	463451	25.0	0.0	2.0	147.5
281	1638137	462197	22.0	8.5	26.0	112.8
284	1634123	459611	2.0	0.0	0.0	114.5
285	1657895	459399	91.5	114.5	161.5	165.5
288	1625201	486167	0.0	3.5	0.0	29.0
289	1615874	483699	58.5	11.0	0.0	52.5
290	1614542	483740	103.5	65.5	117.5	113.5
292	1598661	478505	16.0	58.0	76.5	119.0
293	1599690	481129	61.0	18.5	3.5	80.0
294	1601363	482732	63.0	84.5	101.5	124.0
295	1610566	482514	9.0	9.0	12.5	118.0
296	1611879	481161	5.5	1.5	11.5	73.0
297	1610174	478558	0.0	18.0	0.0	83.5
300	1621988	475327	6.5	52.0	7.5	151.0
301	1613139	475836	151.5	36.0	34.5	149.5
302	1617103	477085	38.0	90.5	15.0	114.5
304	1608176	476958	3.5	53.5	14.5	91.5
306	1609481	476600	11.0	16.0	87.0	136.0
308	1602571	476085	18.0	82.0	15.5	63.5
309	1598626	476177	145.0	113.5	55.5	154.0
310	1597327	477547	7.0	88.0	10.0	89.0
311	1612109	471534	16.5	87.5	26.5	135.0
312	1588233	485531	113.5	112.5	122.0	130.0
313	1581968	487883	28.0	13.5	6.5	176.0
315	1578574	482924	1.5	25.0	7.0	222.0
319	1637587	546548	8.0	25.0	24.5	53.0
320	1637907	545549	29.0	28.5	1.5	34.5
321	1639175	538904	27.5	46.0	5.0	46.0
325	1641780	533625	84.0	87.5	6.5	36.5
326	1652918	530001	10.9	49.0	2.5	12.5
329	1605370	538951	0.0	0.0	7.0	37.0
330	1606022	537936	1.0	24.0	3.0	44.0
331	1599714	531806	16.0	31.5	1.0	33.5
332	1582626	550461	30.5	34.5	22.5	9.0

333	1586626	551754	29.5	31.5	2.5	17.5
334	1569436	551821	4.0	1.5	3.5	29.0
340	1641496	512402	8.0	74.5	2.0	27.0
343	1638644	497928	0.0	0.0	0.0	9.0
346	1617747	511484	78.0	109.5	112.0	72.5
348	1604307	495921	58.0	99.0	61.5	81.5
349	1613456	493020	0.0	0.0	0.0	1.0
350	1583973	518425	77.5	102.5	86.0	82.5
351	1584794	497789	108.5	111.5	131.0	68.0
357	1654281	481877	9.5	25.5	0.0	28.5
359	1609180	478592	7.5	29.5	5.0	33.0
360	1614522	482411	75.5	32.5	9.0	101.5
361	1615845	482374	104.0	47.5	0.5	92.0
362	1589619	488188	20.5	22.0	14.5	127.5
501	1624500	460500	59.0	105.0	12.0	98.0
502	1591000	461500	0.0	10.0	5.0	56.0
503	1596000	468000	55.0	66.0	9.0	157.0
504	1656000	501000	12.0	6.0	10.0	25.0
505	1638500	499000	0.0	0.0	0.0	10.0
506	1642000	487500	0.0	48.0	18.0	35.0
507	1604000	463000	19.0	22.0	7.0	208.0
508	1628500	519000	0.0	15.0	0.0	18.0
509	1594000	514000	5.0	15.0	15.0	80.0
510	1566000	507000	25.0	0.0	2.0	125.0
511	1573500	473500	10.0	20.0	20.0	270.0
512	1643500	472000	42.0	10.0	4.0	60.0
513	1627000	517000	0.0	0.0	0.0	0.0
514	1596000	513500	5.0	10.0	12.0	92.0
515	1654000	495500	7.0	35.0	18.0	45.0
516	1614500	501500	61.0	25.0	18.0	42.0