

# ANALYSIS OF SEDIMENTARY FACIES BY REGIONALIZED CLASSIFICATION

J.C. Davis,<sup>1</sup> J. Harff,<sup>2</sup> W. Lemke,<sup>2</sup> R.A. Olea,<sup>1</sup>  
F. Tauber,<sup>2</sup> G.C. Bohling,<sup>1</sup> Zhou Di<sup>3</sup>

Report of the U.S.-Germany Cooperative Research Project on  
Three-Dimensional Regionalization and Modeling for Sedimentary Basin Analysis  
Supported in part by NSF Grant INT-9111646

KANSAS GEOLOGICAL SURVEY  
Open File Report 95-41

July 1995

---

<sup>1</sup> Mathematical Geology Section, Kansas Geological Survey, The University of Kansas, Lawrence, Kansas, USA.

<sup>2</sup> Baltic Sea Research Institute, Warnemünde, Germany.

<sup>3</sup> South China Sea Institute of Oceanology, Guangzhou, China.

# ANALYSIS OF SEDIMENTARY FACIES BY REGIONALIZED CLASSIFICATION

J.C. Davis,<sup>1</sup> J. Harff,<sup>2</sup> W. Lemke,<sup>2</sup> R.A. Olea,<sup>1</sup>  
F. Tauber,<sup>2</sup> G.C. Bohling,<sup>1</sup> Zhou Di<sup>3</sup>

---

*The spatial pattern of sediment types on the sea bottom in the Darss Sill area of the Baltic Sea can be mapped using regionalized classification of grain size measurements. Unsupervised classification distinguishes seven sediment groups, each representing a sedimentary facies. The genesis of the groups is interpreted from the median grain size and sorting for each group. Six of the groups form a sequence beginning with their source (eroded submarine tills) and extending to depocenters of fine grained sand; a seventh group represents a channel facies. The regional distribution of sediment types is shown on contour maps produced by kriging the probability that each observation is a member of a specified group. A regionalized map of the dominant sediment types on the sea floor in the Darss Sill area is produced by combining these group probabilities. The map confirms assumptions about the dynamics of erosion, transport and deposition within the area. In this application, regionalization works well even though many of the underlying assumptions of the methodology are violated.*

---

**KEY WORDS:** Regionalized classification, geostatistics, marine sediments, grain size distribution, facies type, sea floor mapping.

## INTRODUCTION

Grain-size distributions have been used since the early 1930's to provide clues to the depositional environments in sedimentary basins. The relative proportions of the different sand fractions of clastic sediments commonly are regarded as reliable guides to current velocities and transport energy levels. As an idealization, originally heterogeneous sediments are continuously sorted during transport, and a progression of ever-finer material is deposited in successively less energetic environments as material moves from the sediment source on the basin margin to deep, calm water in the center of the basin. This genetically based model will predict certain patterns of sediment types (facies) within a basin; unfortunately, there are many circumstances that can disrupt this idealized picture and complicate interpretation.

The possible complexities in facies patterns can be assessed by examining the distribution of sediments in modern depositional basins, where the environments are apparent and inference is not necessary. The Darss Sill area within the Baltic Sea (Fig. 1) provides an example where the primary source of sediments is not by transport from the basin margin; rather, sediments are derived from within the basin itself. Since the classical genetic model of facies patterns is not appropriate for such a circumstance, we have turned instead to a statistical procedure which constructs a pattern of sedimentary facies based directly on the observations.

The Darss Sill is a bottleneck in the exchange of water between the North Sea and the deep basins of the Baltic Sea. The dynamics of water exchange during the Postglacial geologic history of the Baltic Sea have resulted in a spatial distribution of sediments that have been investigated for many years by the Baltic Sea Research Institute-Warnemünde (Lemke, 1992) and by the Danish

---

<sup>1</sup> Mathematical Geology Section, Kansas Geological Survey, The University of Kansas, Lawrence, Kansas, USA.

<sup>2</sup> Baltic Sea Research Institute, Warnemünde, Germany.

<sup>3</sup> South China Sea Institute of Oceanology, Guangzhou, China.

## Analysis of Sedimentary Facies by Regionalized Classification

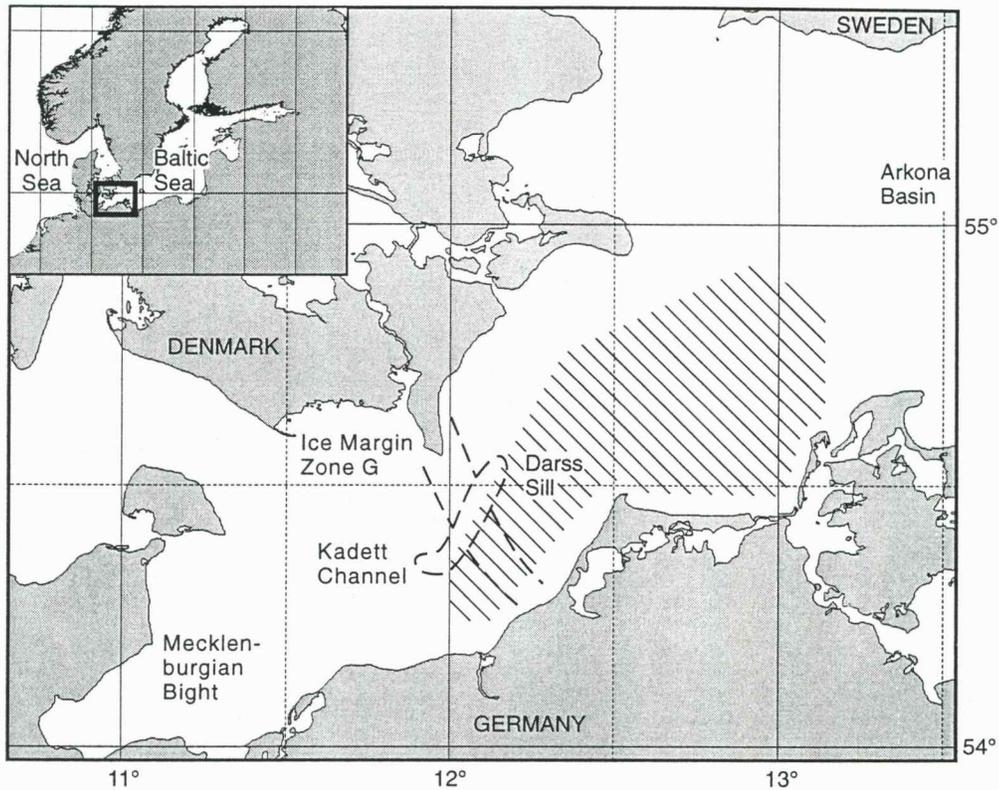


Figure 1. Darss Sill and surrounding area of the eastern part of the Baltic Sea.

Geological Survey (Lemke and others, 1994). Grain size measurements from almost 1,300 bottom samples collected in the Darss Sill area were analyzed using the method of regionalized classification (Harff and Davis, 1990). The objective was to determine the distribution of sediments in the area of investigation and to produce a genetic interpretation of the resulting pattern.

### THE DARSS SILL AREA

The Baltic Sea is the world's largest modern brackish water reservoir. It is the largest body of water in western Europe, and is connected with the North Sea by the Danish straits. Because of the humid climate, the Baltic Sea is characterized by a positive balance of runoff water into the North Sea. Marine water flows into the Baltic, forcing oxygen-rich ocean water into the deep basins of the central Baltic. These currents flow past the Darss Sill, which forms a natural barrier between the Mecklenburgian Bight in the west and the Arkona Basin in the east. This barrier was formed during the Late Weichselian glaciation (14,000 to 13,000 yrs BP) at the marginal zone of the "G" advance and consists of glacial tills which now crop out on the sea bottom. Postglacial drainage created the Kadett Channel which was incised into the tills. This channel divides the Darss Sill into a northern (Danish) part and a southern (German) part, each characterized by different current systems. The currents move in a counter-clockwise direction because of coriolis forces, with inflowing ocean water passing through the Kadett Channel and over the southeastern part of the Darss Sill while outflowing water from the northeast follows the northwestern border of the Baltic Basin. It passes over the northern part of the Darss Sill before leaving the Belt Sea through the Great Belt.

The glacial tills of the Darss Sill are covered by a thin layer of lag sediments which are the main source of sediments within this part of the Baltic (Fig. 2). Because of the dominant current

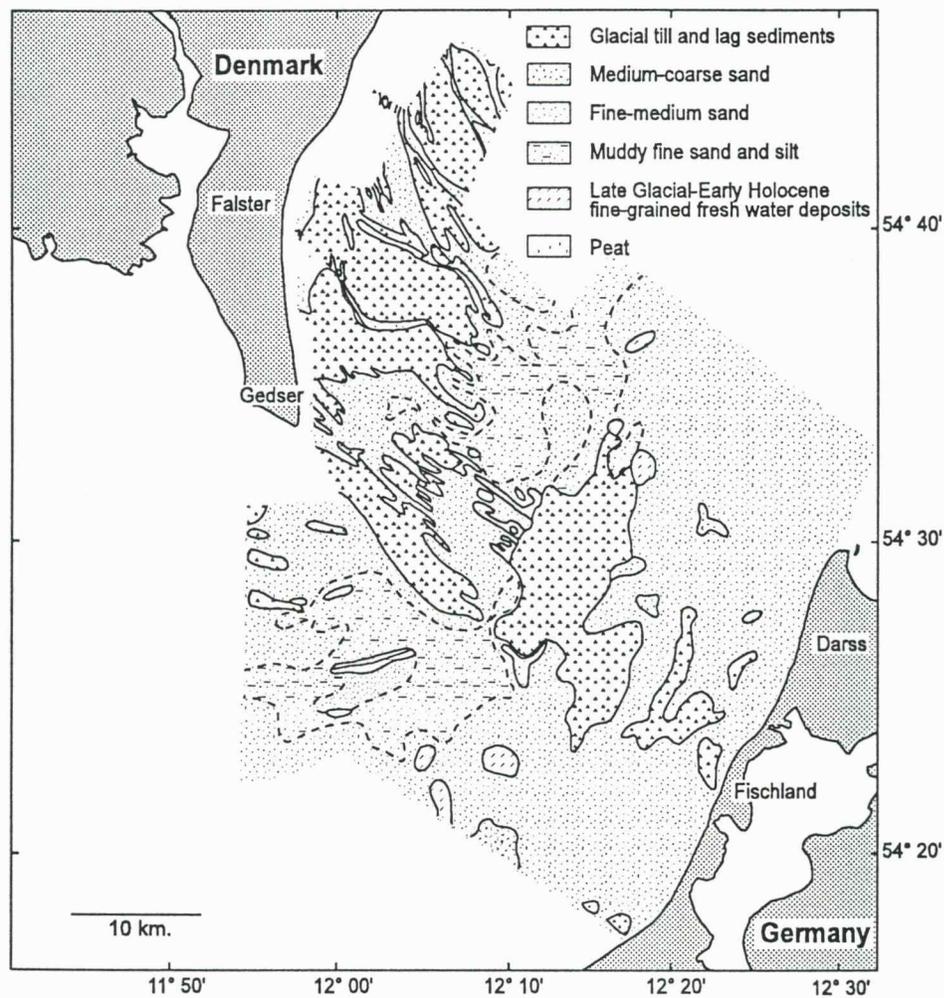


Figure 2. Generalized geologic map showing bottom sediments in the Darss Sill area, Baltic Sea. After Lemke and others, 1994.

directions, sediments in the German part of the Baltic are transported from the southeast part of the area to the northeast and east, becoming progressively better sorted. This is in contrast to the Danish part of the Baltic, where the main transport direction is to the southwest and west (Fig. 3). North of the eastern tip of the Zingst Peninsula, glacial tills crop out at some locations.

### BALTIC SEDIMENT DATA

The data represent 1281 bottom samples collected by grab sampler along marine traverses conducted by the oceanographic research vessels R.S. Alexander von Humboldt and R.S. Prof. Albrecht Penck. The UTM coordinates of each grab sample location are a part of the data. The sediments were dried and separated by mechanical sieving into eight size fractions ranging from less than  $63\mu\text{m}$  (silt and finer) to over  $2000\mu\text{m}$  (gravel and coarser). The amount of material retained by each sieve was recorded as the percent of the total sample weight. Table 1 gives the variable names, grain size class limits, number of observations, average weight percent, and standard deviation of the weight percent of each size fraction.

*Analysis of Sedimentary Facies by Regionalized Classification*

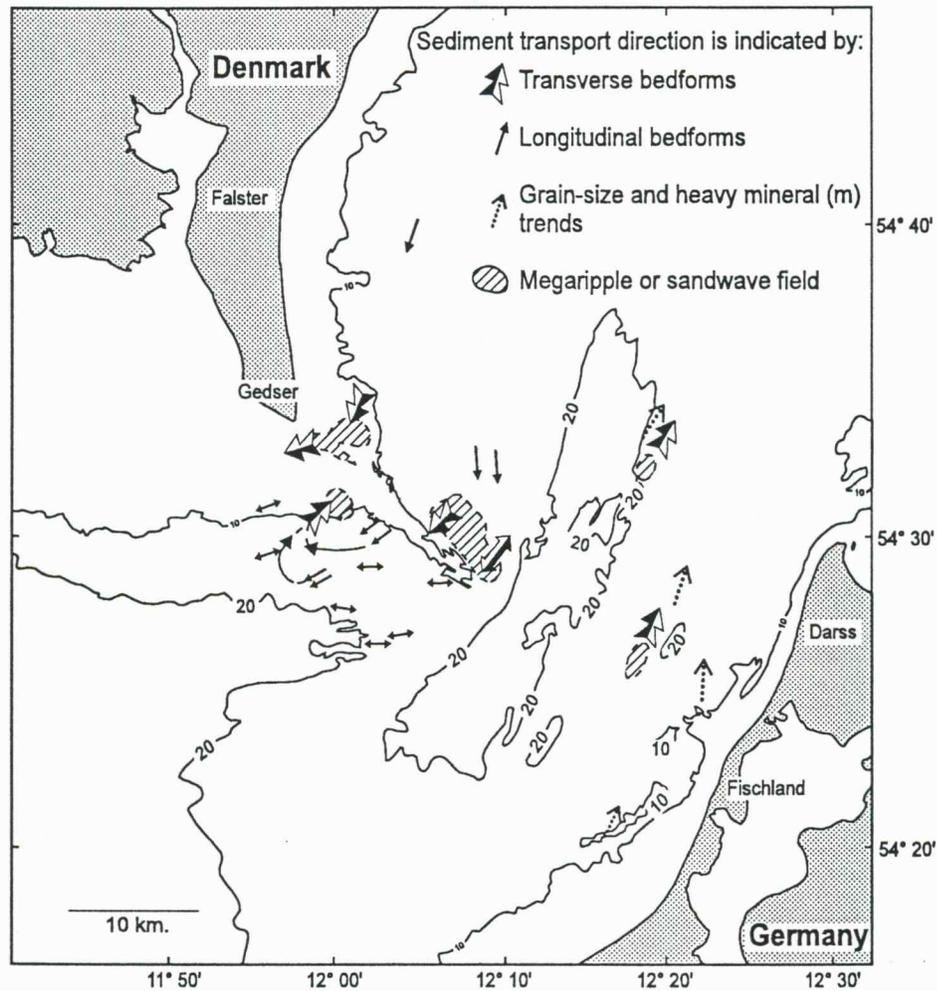


Figure 3. Sediment transport directions in the Darss Sill area, Baltic Sea. After Lemke and others, 1994.

TABLE 1. VARIABLES USED FOR REGIONALIZATION OF BOTTOM SEDIMENTS IN THE GERMAN PART OF THE DARSS SILL AREA OF THE BALTIC SEA

Variable	Grain size class limits	No. of observations	Average wt. percent	Std. dev. wt. percent
Gravel	> 2000 $\mu$ m	125	5.6	6.8
Very coarse sand	2000-1000 $\mu$ m	356	4.8	5.0
Coarse sand	1000-630 $\mu$ m	682	5.4	6.9
Medium sand	630-400 $\mu$ m	1186	7.0	9.7
Medium fine sand	400-200 $\mu$ m	1269	35.3	24.9
Fine sand	200-100 $\mu$ m	1281	44.1	38.5
Very fine sand	63-100 $\mu$ m	1273	6.8	9.4
Silt	< 63 $\mu$ m	1226	3.1	8.4

Such data have several characteristics that complicate their analysis. Conventional measures of size express dimensions of individual objects such as grains directly, using a ratio scale of measurement (*i.e.*, the lengths of grain axes in millimeters or grain diameters in microns). In contrast, sediment grain size analyses express size indirectly, as the relative weight of all grains that pass through the openings of one sieve but are retained by the next smaller sieve. The number of grains involved is not known. Since the measurements do not relate to individual observations,  $n$  is unknown and conventional statistics (mean, variance) cannot be computed. The quantities referred to as “grain size statistics” are not true statistics, but are graphical descriptors of the shape of the histograms that express the weight percent in each grain size class. The unique characteristics of grain size measurements have been discussed at length by Griffiths (1967).

A more serious complication arises because the weights of individual size fractions are expressed in relative form as the percent of the total sample weight. Such data constitute a *composition*, a set of variables which sum to a constant. This constraint restricts the range of possible values and induces negative relationships between variables, as one constituent cannot increase unless others decrease an equivalent amount. The treatment of compositional variables has been examined in detail by Aitchison (1986). The effects of the nature of the grain size data on computations used in regionalization will be discussed in a later section.

#### REGIONALIZED CLASSIFICATION PROCEDURE

The concept of multivariate classification of “geological objects” developed by Voronin (1967) and extended by Rodionov (1981) has been combined with elements of regionalized variable theory by Harff and Davis (1990) to produce what is called “regionalized classification.” The mathematical formalism will not be repeated here; instead, the computational aspects will be summarized and certain underlying assumptions highlighted. The objective of regionalized classification is to subdivide a two- or three-dimensional portion of the earth’s crust into contiguous parts called regions that are as internally homogeneous as possible and as distinct as possible from adjacent regions (Harff, Davis, and Eiserbeck, 1993). In this application, the area to be regionalized is the Darss Sill study area of the Baltic Basin, and the properties on which the regions are based are grain size measurements from bottom samples. In conventional geological terms, the regions correspond to sedimentary facies.

The initial step is typification, in which the observations are subdivided into groups based on their mutual similarities. An unsupervised hierarchical clustering procedure such as Ward’s algorithm, which uses a within-cluster minimum variance criterion, can be used to produce candidate groups (Bock, 1974). In common with other Q-mode procedures, there is no theoretical underpinning from which to estimate the appropriate number of groups or to evaluate their significance (Davis, 1986). External criteria may prove useful, and the degree of contiguity of cluster members also may be indicative. At the highest level of clustering, all observations belong to a single cluster and of necessity form a contiguous group. At the next lower level, the single cluster is split into two clusters whose members ideally occupy two distinct, separate areas. Because the spatial coordinates of the observations are not used in the clustering process, there is no guarantee that this will occur—it is possible that members of the two clusters will be spatially intermixed. In such a circumstance, we must conclude that regions either do not exist, or the variables chosen to not reflect the regionalization. Typically, when the number of clusters is limited, the cluster members will form a pattern consisting of approximately the same number of (or slightly more) discrete areas. As the number of clusters increases, however, the spatial organization breaks down and the resulting map is a chaotic mixture of individual points. The level in the hierarchy at which this occurs provides a clue to the appropriate number of regions.

In the Darss Sill study, the appropriate number of clusters was chosen by examining the map patterns formed by cluster members, and by comparison with conventional maps of bottom sediments. It was decided that a regionalization into seven classes would be most effective.

### *Analysis of Sedimentary Facies by Regionalized Classification*

For each cluster, the group centroid and within-group covariance matrix (actually, the matrix of sums of squares and cross products) are calculated and the individual within-group matrices are pooled to form matrix **E**. The between-group covariance matrix **H** also is calculated (again, the matrix of sums of squares and cross products is used in practice). Because grain-size measurements are compositional data, these matrices have troublesome properties that will be considered later. The eigenvectors of  $\mathbf{E}^{-1}\mathbf{H}$  are linear discriminant functions, orthogonal axes onto which the observations can be projected (Fisher, 1936). To classify individual observations, it is necessary to calculate the similarity between each observation and the centroids of the groups to which it might be assigned. The similarity metric used is Mahalanobis' distance, which takes into account the relative inflation of the groups about their centroids. For classification purposes, these multivariate distances can be turned into a set of posterior probabilities that the observation belongs to each respective group. The observation is assigned to the group for which the probability of membership is the greatest (Tatsuoka, 1971).

Grain size measurements such as the percentage of fine sand in a sample can be considered to be a spatially varying stochastic variable whose spatial continuity can be expressed by its semivariogram; in other words, a regionalized variable (Journel and Huijbregts, 1978). Since discriminant functions are linear combinations, the discriminant scores based on grain size classes also are regionalized variables, as are the Mahalanobis' distances and the group membership probabilities. At every sample locality, the probability that the observation is a member of group *k* can be determined, and from these, a semivariogram describing the spatial continuity of the group *k* probability function can be estimated. Using parameters from an appropriate model fitted to the semivariogram, kriging estimates of the probability of classification in group *k* can be made at locations where no observations are available; it is convenient to make such estimates at the nodes of a regular grid covering the study area (Deutsch and Journel, 1992). The process can be repeated for all *k* groups, resulting in a series of probability surfaces that describe the likelihood of group membership at every grid location in the area. A specific grid node is assigned membership in the group for which the probability is the highest.

The final phase in regionalization is to produce a grid showing the maximum probability of assignment to any group. As a practical matter, no group assignment is made at grid cells where the highest probability is less than 0.5; this distinguishes those areas that can be placed in a group with reasonable certainty from areas whose classification is unclear. The latter mostly form boundaries between regions but may also delineate areas which do not fit well into the system of classes that has been specified. A map showing the membership assignment at each grid node is the final expression of the regionalization. The regionalization map should be accompanied by a contour map of the maximum probability of assignment, because this expresses the reliability of the regions.

### TYPIFICATION OF DARSS SILL SEDIMENTS

In the process of regionalization, all observations in the Darss Sill study were assigned to seven groups. These groups are represented by a "typical" member represented by the compositional centroids of the groups. Table 2 summarizes the typical mean weight percents in the size classes of each group. The groups were ordered and numbered by decreasing mean grain size. Although not used for regionalization, the graphic median grain size and the graphic sorting also are included in the table; these are conventional summary measures of the grain size distribution. Figure 4 is a cross-plot of the graphic median versus graphic sorting, with the group membership of the observations indicated by colors assigned to the regions in the final regionalized map.

The process of hierarchical agglomerative clustering is shown as a dendrogram in Figure 5. Although a complete hierarchical clustering was performed of all 1831 observations, only the highest seven agglomerated groups are shown. The histograms on the dendrogram represent the average compositions of the group members at each stage of clustering. The histogram bars at the lowest

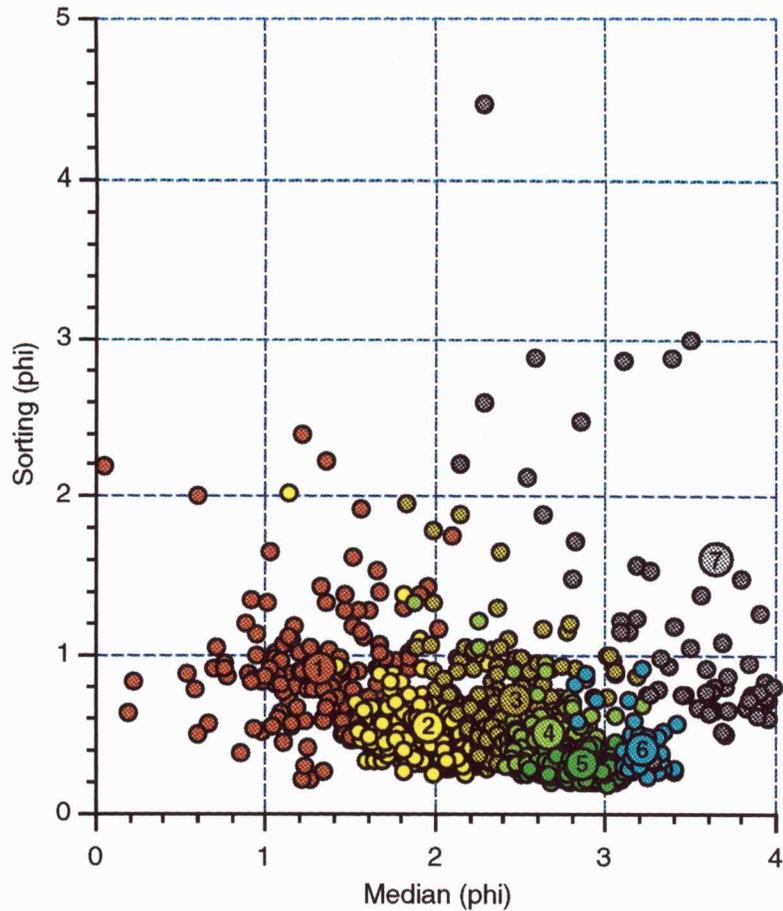


Figure 4. Median grain size versus sorting of bottom sediments from the Darss Sill area, Baltic Sea. Statistics are calculated from the shape of weight-percent size cumulative curves.

level shown correspond to the average class interval weight percents given in Table 2. The color convention is the same as used in Figure 4.

Following the initial typification, the probabilities of class membership can be determined by computing Mahalanobis' distance between every observation and the centroids of the seven groups. Mahalanobis' distance reflects the relative inflation of the group as well as the differences between the observation and the group centroid, and can be found by linear discriminant analysis. This step may result in some reallocation of observations to different clusters to which they have a higher probability of membership than to the original clusters. Table 3 is a reclassification table for the Darss Sill observations.

A genetic interpretation can be assigned to the results of sediment classification. The poorly sorted Group 1, with an unusually high proportion of gravel and coarse sand, represents the remnants from submarine erosion of a glacial till. Group 2 also is erosional in origin, but shows a transitional character. Group 3 belongs to a depositional facies but its relatively poor sorting suggests that it is a transitional sediment. The more typical depositional facies include Group 4 which has transitional characteristics, Group 5 which is typical of depocenters, and Group 6, which represents a sediment that has bypassed the depocenter and has been deposited in a distal position. Group 7 represents a channel environment where erosion, transportation, and deposition change very rapidly.

Computations were performed using routines from SAS, JMP, and MathCAD.

*Analysis of Sedimentary Facies by Regionalized Classification*

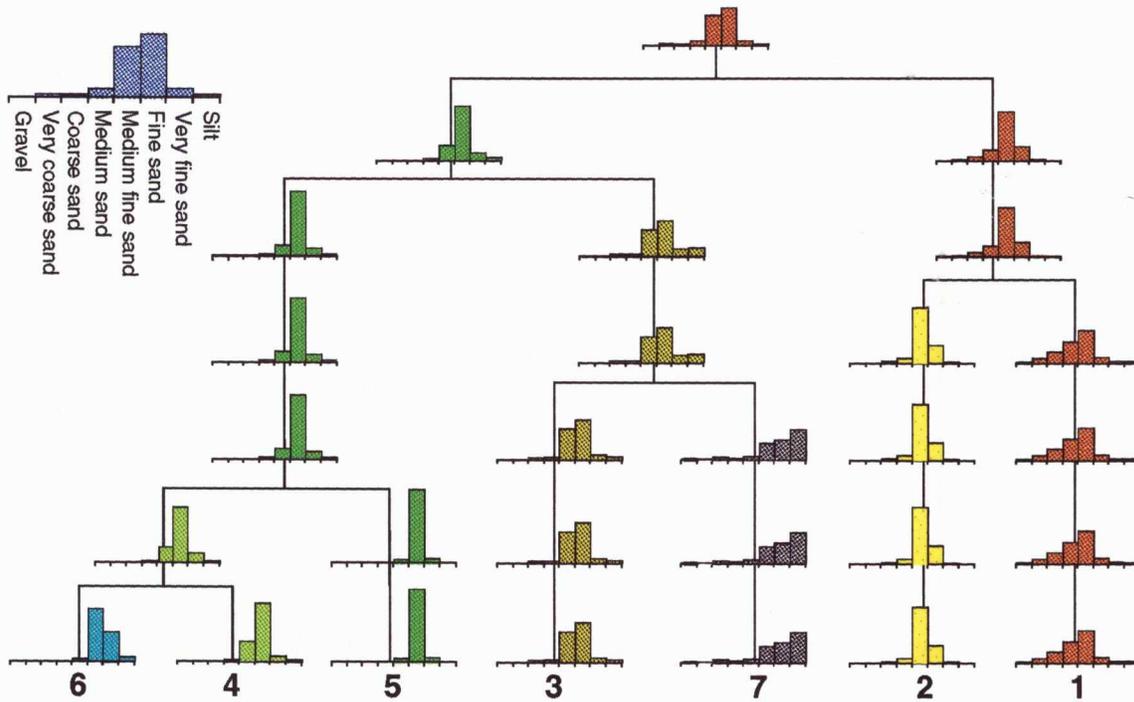


Figure 5. Dendrogram showing seven highest levels of agglomeration of sediment data from the Darss Sill area, Baltic Sea.

TABLE 2. GRAIN SIZE DISTRIBUTIONS CORRESPONDING TO THE CENTROIDS OF THE SEVEN GROUPS DEFINED BY HIERARCHICAL CLUSTERING OF DARSS SILL DATA

Group	<i>n</i>	Very coarse sand	Coarse sand	Med. sand	Med. fine sand	Fine sand	Very fine sand	Silt
1	167	2.3	7.1	14.3	26.4	39.0	8.3	1.6
2	396	0.3	0.7	2.1	7.1	64.7	21.4	3.1
3	209	0.4	0.5	1.2	3.2	37.0	46.7	6.8
4	232	0.1	0.3	0.8	1.9	22.7	66.1	6.2
5	224	0.0	0.0	0.1	0.5	5.6	86.2	6.7
6	56	0.0	0.0	0.0	0.3	1.9	59.5	34.7
7	52	1.9	2.0	3.7	3.9	6.0	20.3	25.3

**REGIONALIZATION OF THE DARSS SILL USING GRAIN SIZE DATA**

As an initial step, a geostatistical study was made of the bathymetry of the Darss Sill area. Depth measurements are free of the numeric complications that affect grain size data and can be analyzed with little difficulty. Since the bottom topography may have a profound effect on the distribution of sediment types, computing the semivariogram of water depth and producing a bathymetric map seems a prudent initial step in the regionalization process. Because water depth is an anisotropic

TABLE 3. RESULTS OF SUPERVISED RECLASSIFICATION. PRIOR ASSIGNMENTS OF OBSERVATIONS TO CLUSTERS ARE COMPARED TO POSTERIOR GROUP ASSIGNMENTS BASED ON PROBABILITIES OF GROUP MEMBERSHIPS FROM DISCRIMINANT ANALYSIS

Prior	1	2	3	4	5	6	7	$p_i$
Posterior								
1	0.938	0.034	0.000	0.000	0.000	0.000	0.064	0.1304
2	0.062	0.879	0.000	0.000	0.000	0.000	0.000	0.2693
3	0.000	0.087	0.862	0.045	0.000	0.000	0.000	0.1655
4	0.000	0.000	0.051	0.910	0.005	0.000	0.000	0.1663
5	0.000	0.000	0.000	0.032	0.995	0.018	0.000	0.1764
6	0.000	0.000	0.010	0.009	0.000	0.982	0.000	0.0453
7	0.000	0.000	0.076	0.005	0.000	0.000	0.936	0.1304
Sum								1.0000
Count	161	381	196	222	219	55	47	1831

property in the study area, the sea bottom northeast of the Darsser Ort was modeled separately from the sea bottom in the western part of the area. Figure 6 shows the experimental semivariograms and fitted models for the two areas. Directional semivariograms oriented perpendicular to the regional sea floor slope were used to avoid the nonstationarity caused by drift. The semivariograms show that sea floor topography is rougher in the western part, where the range of the fitted Gaussian model is only 8.32 km. In the eastern part, the Gaussian model has a range of 11.72 km. A bathymetric map produced by kriging based on these two models is shown in Figure 7. The map clearly shows the Kadet Channel and some smaller channels in the western part of the area, and also the smooth change in basinal structures in the eastern part.

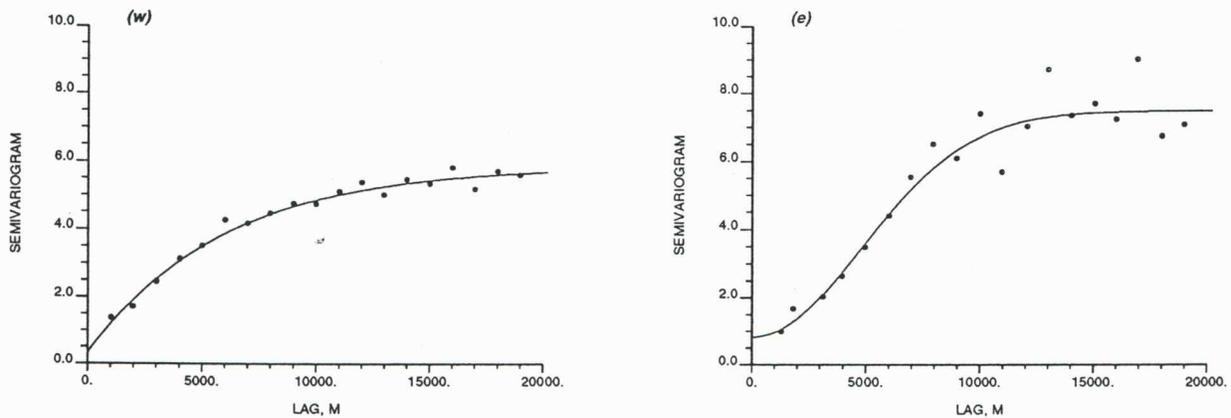
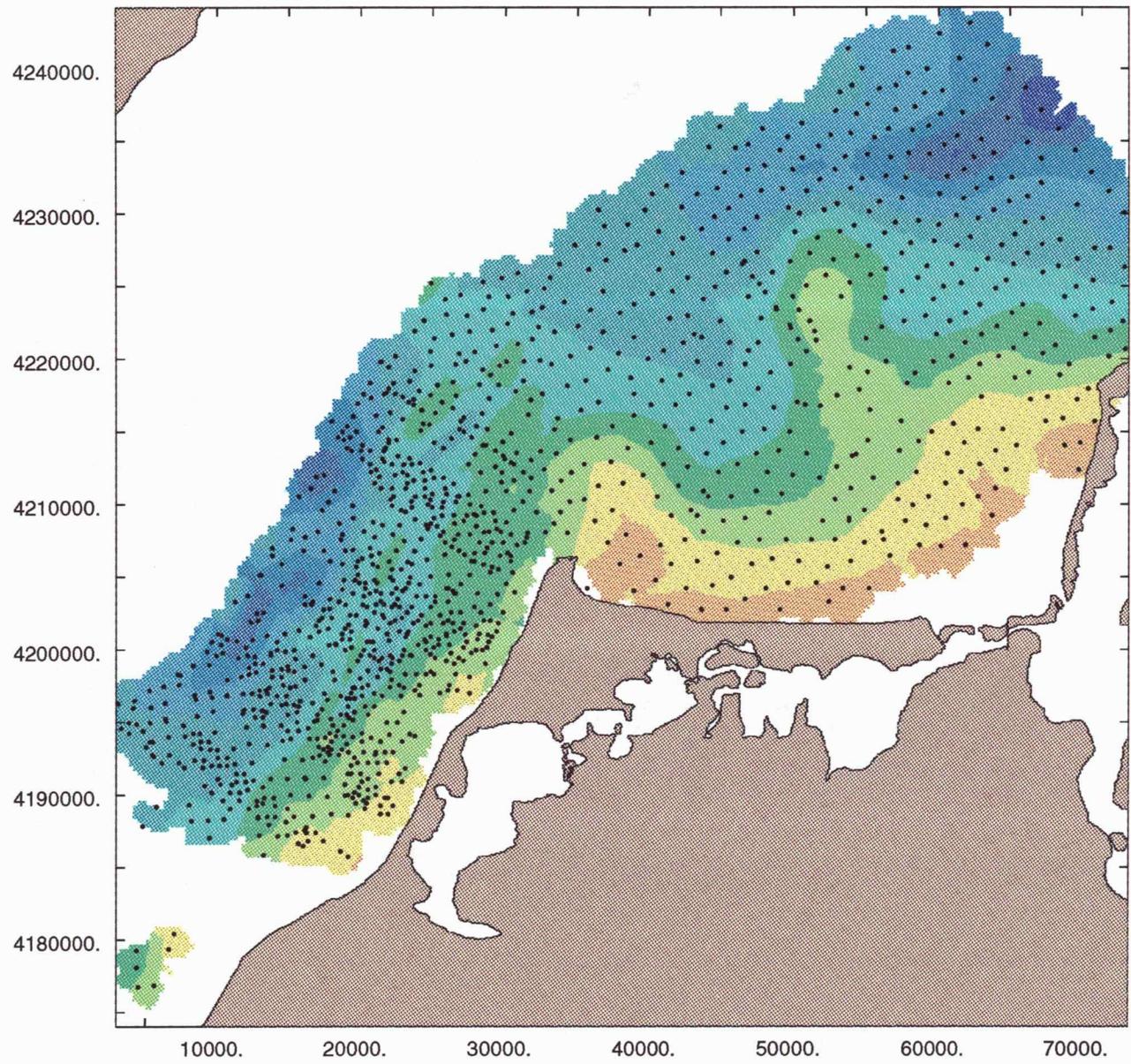
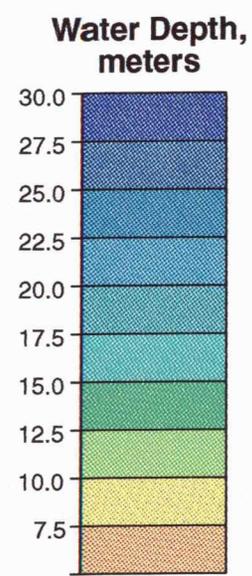


Figure 6. Experimental semivariograms and fitted models for water depth at sampling points in (e), the eastern part and (w), the western part of the Darss Sill area, Baltic Sea. Y-axes in square meters, X-axes in meters. Model for eastern part is Gaussian with a nugget of 1.186, sill of 3.7, and range of 8317.3. Model for western part is Gaussian with a nugget of 0.8, sill of 6.7, and range of 11,718.6.

To perform regionalization of sediment types, it is necessary to map the probabilities of group membership. Since the probabilities are calculated as linear combinations of the grain size measurements, the probabilities can be regarded as regionalized variables. Figure 8 shows experimental



**Figure 7.**  
**Bathymetric map of the Darss Sill area, Baltic Sea. Kriging based on models in Fig. 6.**



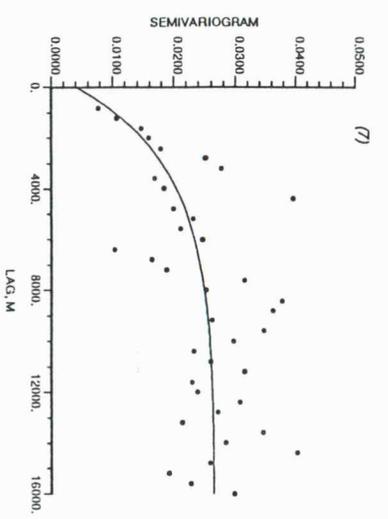
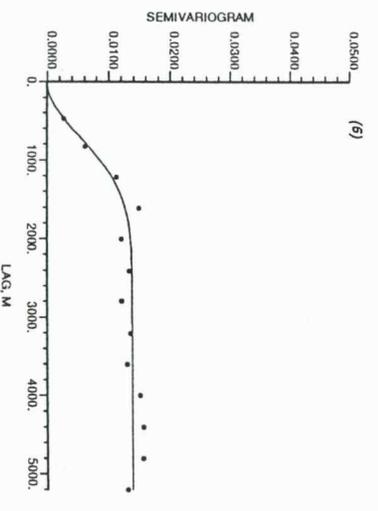
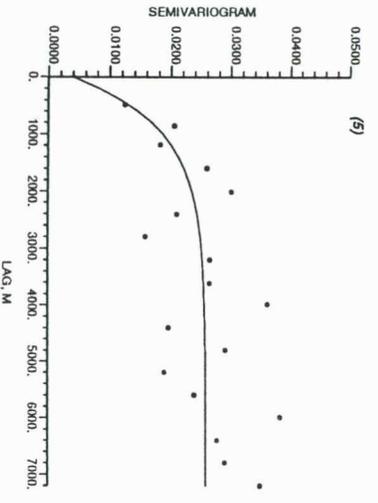
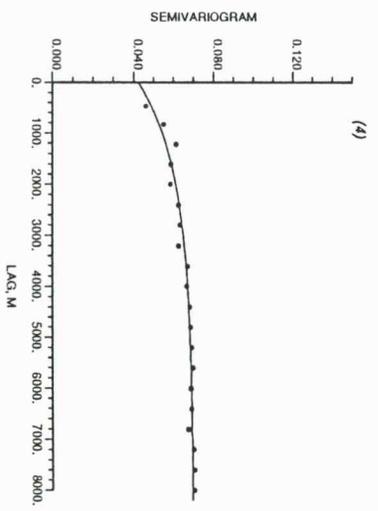
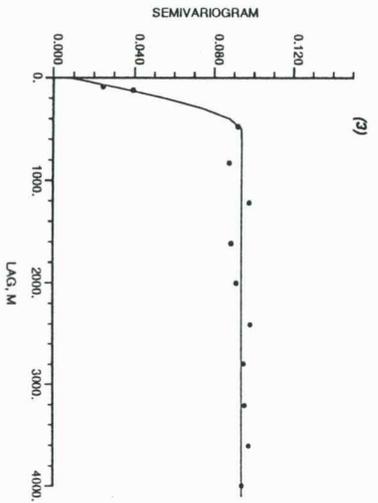
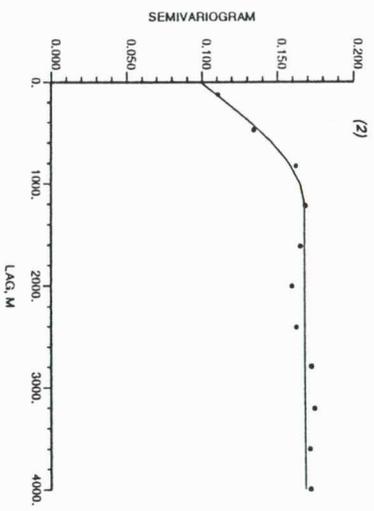
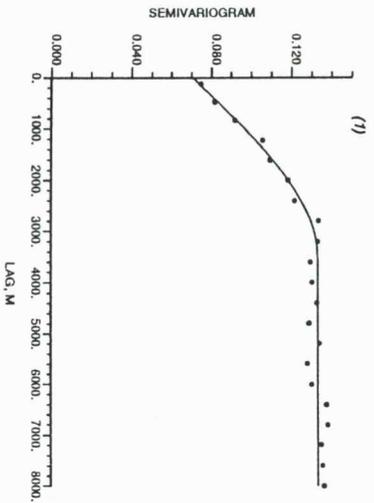
semivariograms for each group probability and the semivariogram used to fit these data. Based on these semivariogram models, kriging was performed and the probabilities of group membership at the nodes of a regular spatial grid were estimated. This yields seven grids of estimates, each of which can be displayed as a contour map. An individual map indicates locations which are most likely to belong to a specific grain size group and distinguishes high-probability areas (shown in shades of red) from those where the probability of group membership is low (shown in shades of blue). Comparison of the seven maps shows that the high-probability areas of different groups do not overlap; each map defines separate areas that have a high probability of belonging to a single specific group.

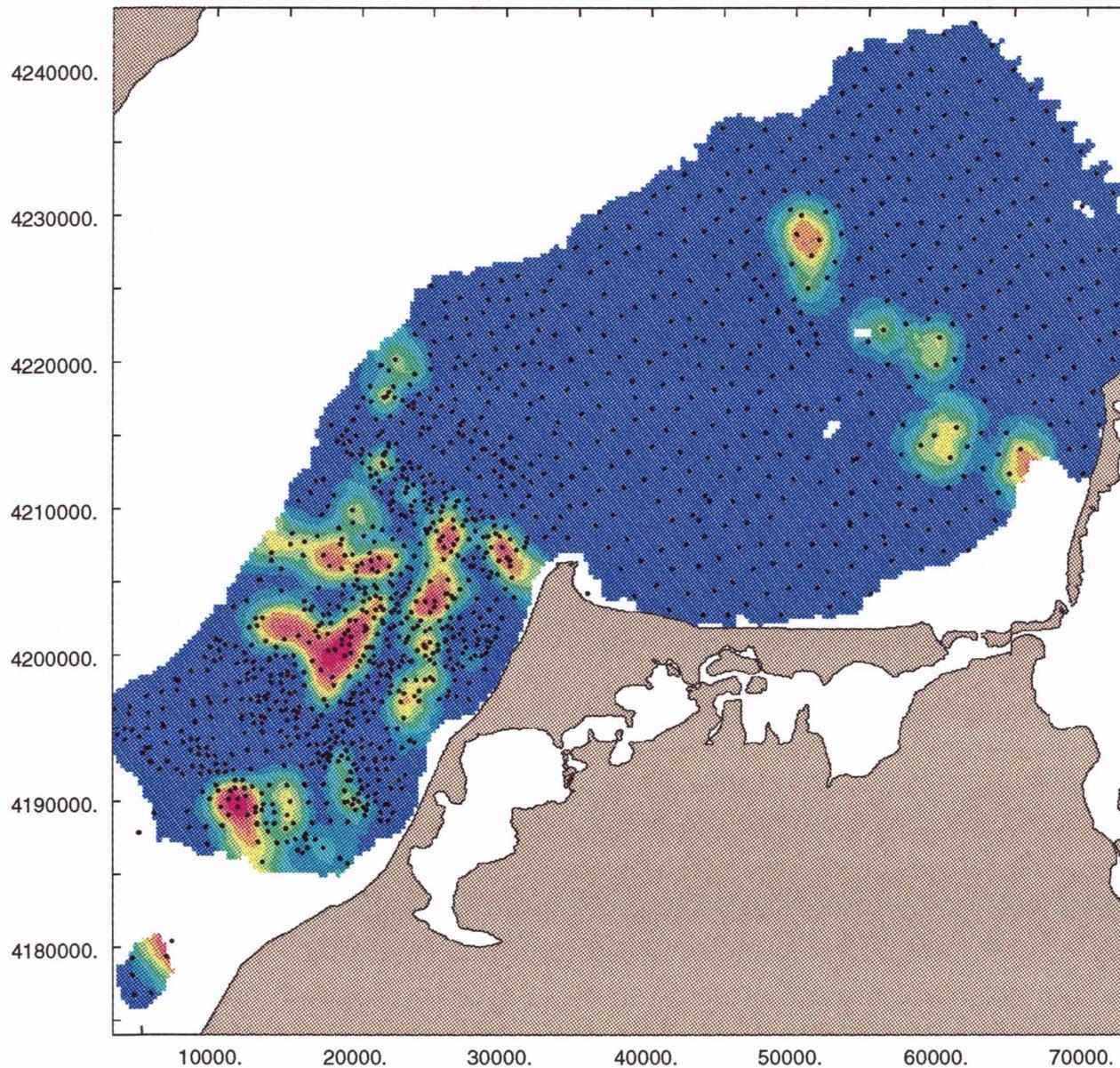
The probability of membership in Group 1 is shown in Figure 9. The areas of high probability are close to the submarine outcrops of glacial till. Two source areas, one in the western and another in the eastern part of the study area, are clearly recognizable. Figure 10 shows the probability of membership in Group 2, the next group in the genetic sequence. The link of the region of high probability with the sediment source area and the source of lag sediments is obvious. A ring-like pattern in the eastern part of the map may reflect the circular current field. Sediments of Group 3 also are transitional types but are more closely related to their depositional environments. Figure 11 shows the probability of membership in Group 3. The generally lower probabilities correspond to the probabilities shown in Table 2, reflecting the transitional nature of this class. The ring-like feature in the east is clearly visible. Comparable patterns appear in Figure 12, showing the probability of membership in Group 4. This similarity in pattern tends to confirm that the genetic sequence of the groups is a response to sediment fractionation during transport. Regions having high probabilities for membership in Group 5 represent the depocenters in the study area. Figure 13 clearly defines the main depositional centers in the eastern area. Two are apparent; one at the western border of the region and one in the center of the ring-like feature apparent in Figures 9 and 10. The distal concentration of Group 6 beyond the western depocenter is shown on Figure 14. In Figure 15, the special nature of Group 7 is shown; it is confined to the Kadet Channel and reflects the unique transport conditions within this erosional form.

The interpretations derived from patterns in Figures 9 through 15 are summarized in Figure 16, which shows the regionalization as a form of geological map. This map was created by allocating each grid cell to the group for which its membership probability is the highest. The reliability of the resulting regionalization can be judged from Figure 17, which shows the maximum probability of membership at every point. As we might expect, areas of transition between the sources of sediments and sediment depocenters are characterized by relatively low probabilities of correct classification. If we construct a regionalization based only on classification probabilities that exceed 0.5, we produce the map shown in Figure 18. Areas where the probabilities of correct classification are relatively high are shown in color. Zones where the proper classification is uncertain are indicated by a gray tone; these represent environments of transition, particularly between source areas and different depocenters.

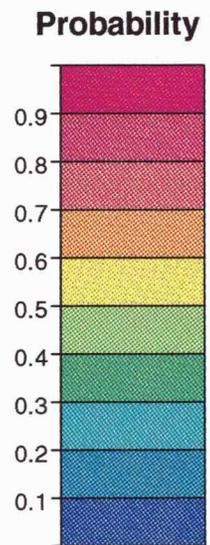
**Figure 8.** Experimental semivariograms and fitted models for Darss Sill area of the Baltic Sea. Variables are the probabilities of classification of observations into Groups 1 through 7, estimated at each sampling locality. Y-axes in probabilities squared; X-axes in meters. Model for Group 1 is Gaussian, with nugget of 0.078, sill of 0.055, and range of 2853.4. For Group 2, model is spherical, with nugget of 0.1, sill of 0.066, and range of 1238.8. For Group 3, model is spherical, with nugget of 0.02, sill of 0.08, and range of 570.3. For Group 4, model is exponential, with nugget 0.045, sill of 0.03, and range of 4404.8. For Group 5, model is Gaussian, with nugget of 0.015, sill of 0.02, and range of 7000.0. For Group 6, model is Gaussian, with sill of 0.012 and range of 1725. For Group 7, model is exponential, with nugget of 0.004, sill of 0.022, and range of 8862.6.

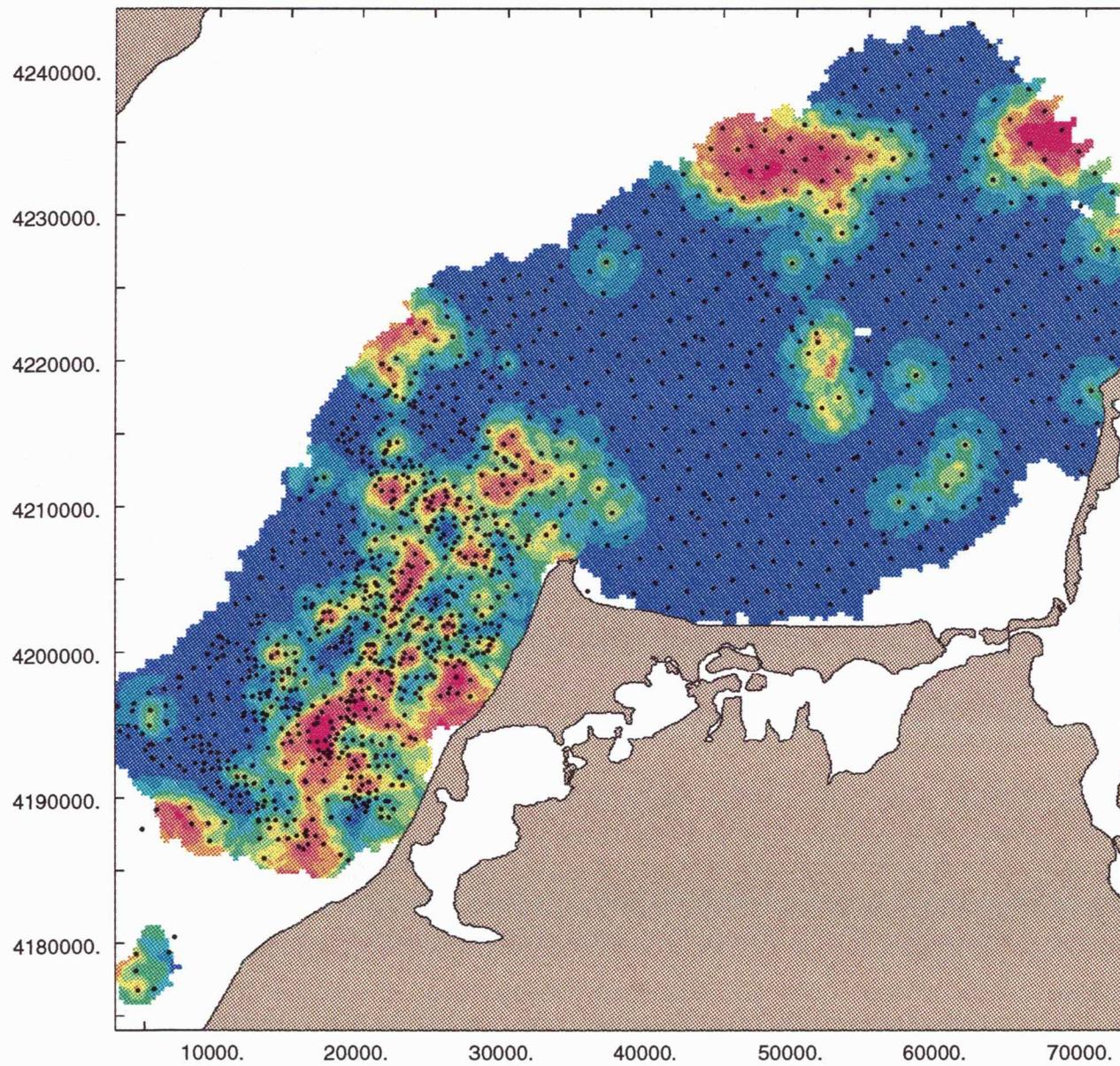
*Analysis of Sedimentary Facies by Regionalized Classification*



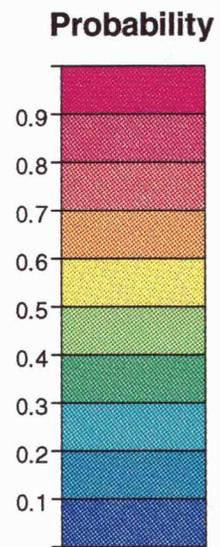


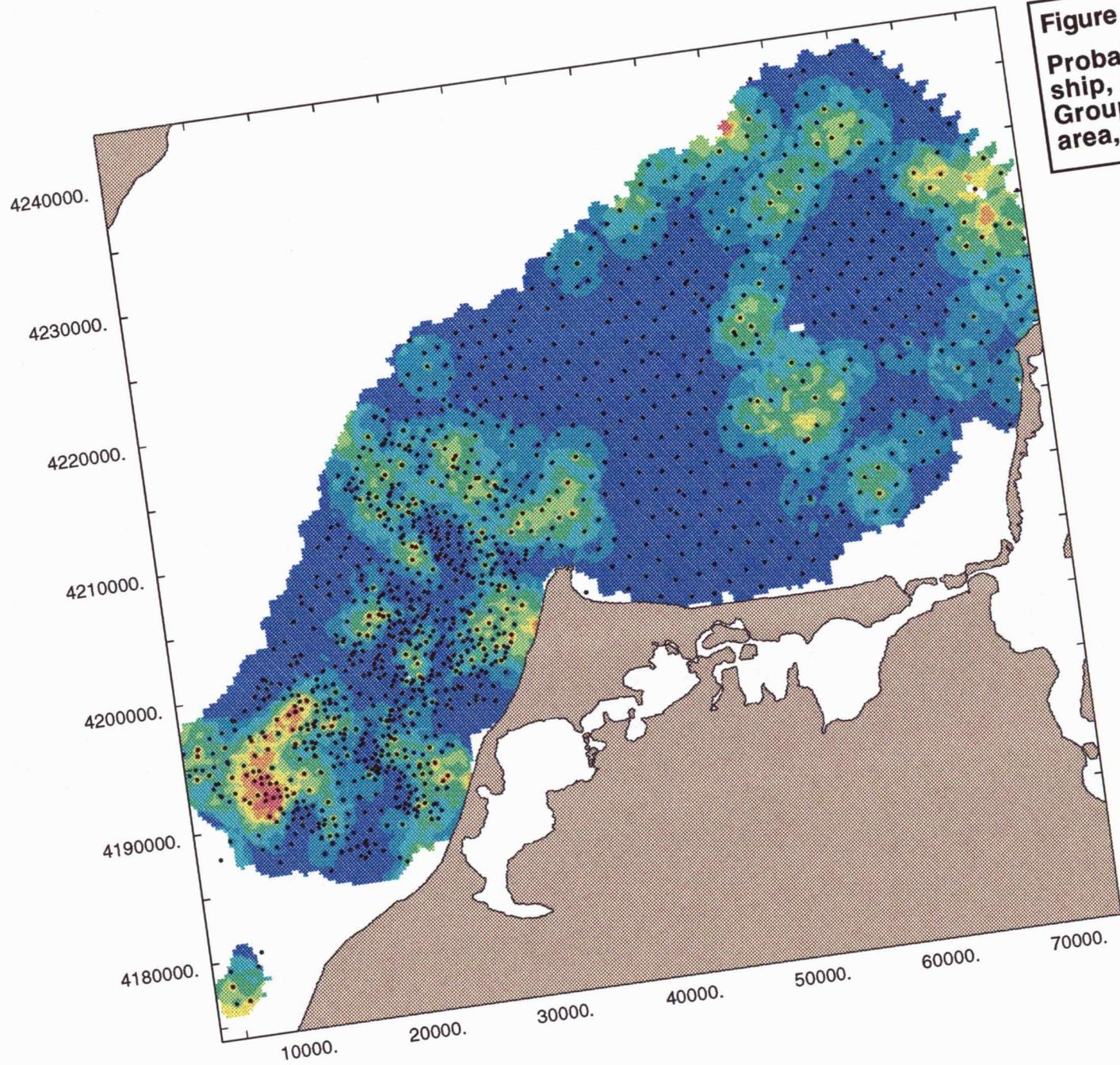
**Figure 9.**  
**Probability of membership,**  
**bottom sediment**  
**Group 1, Darss Sill**  
**area, Baltic Sea.**



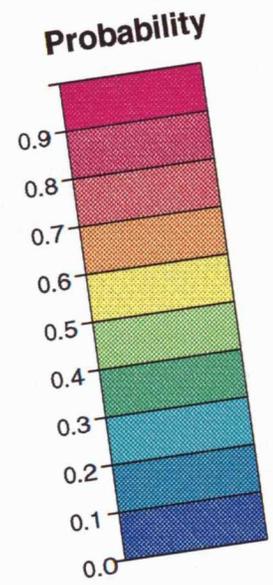


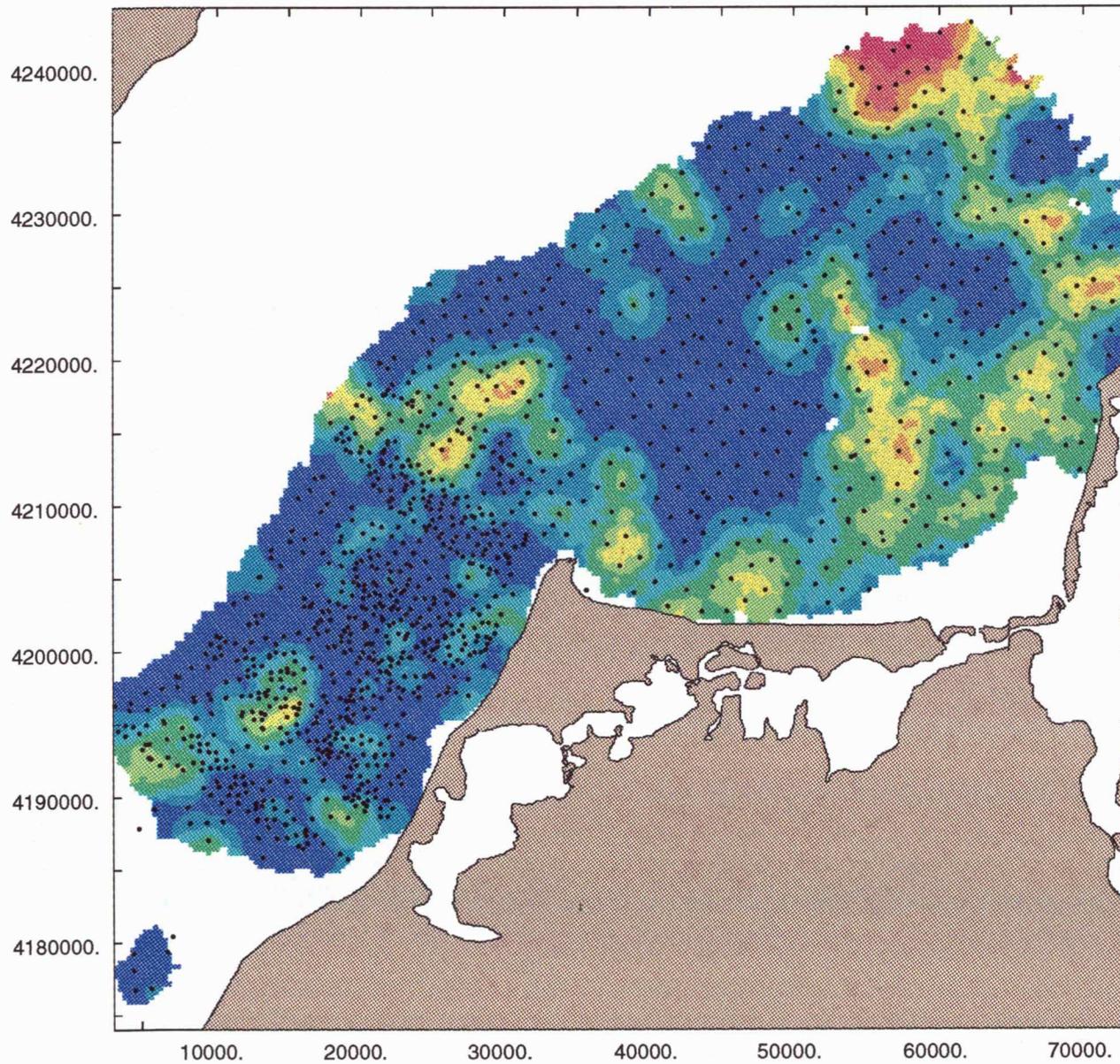
**Figure 10.**  
**Probability of membership,**  
**bottom sediment**  
**Group 2, Darss Sill**  
**area, Baltic Sea.**



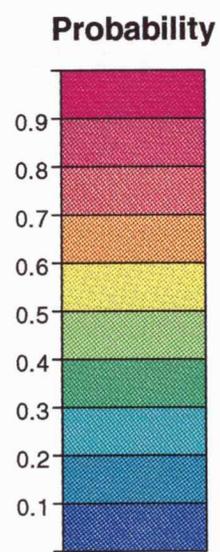


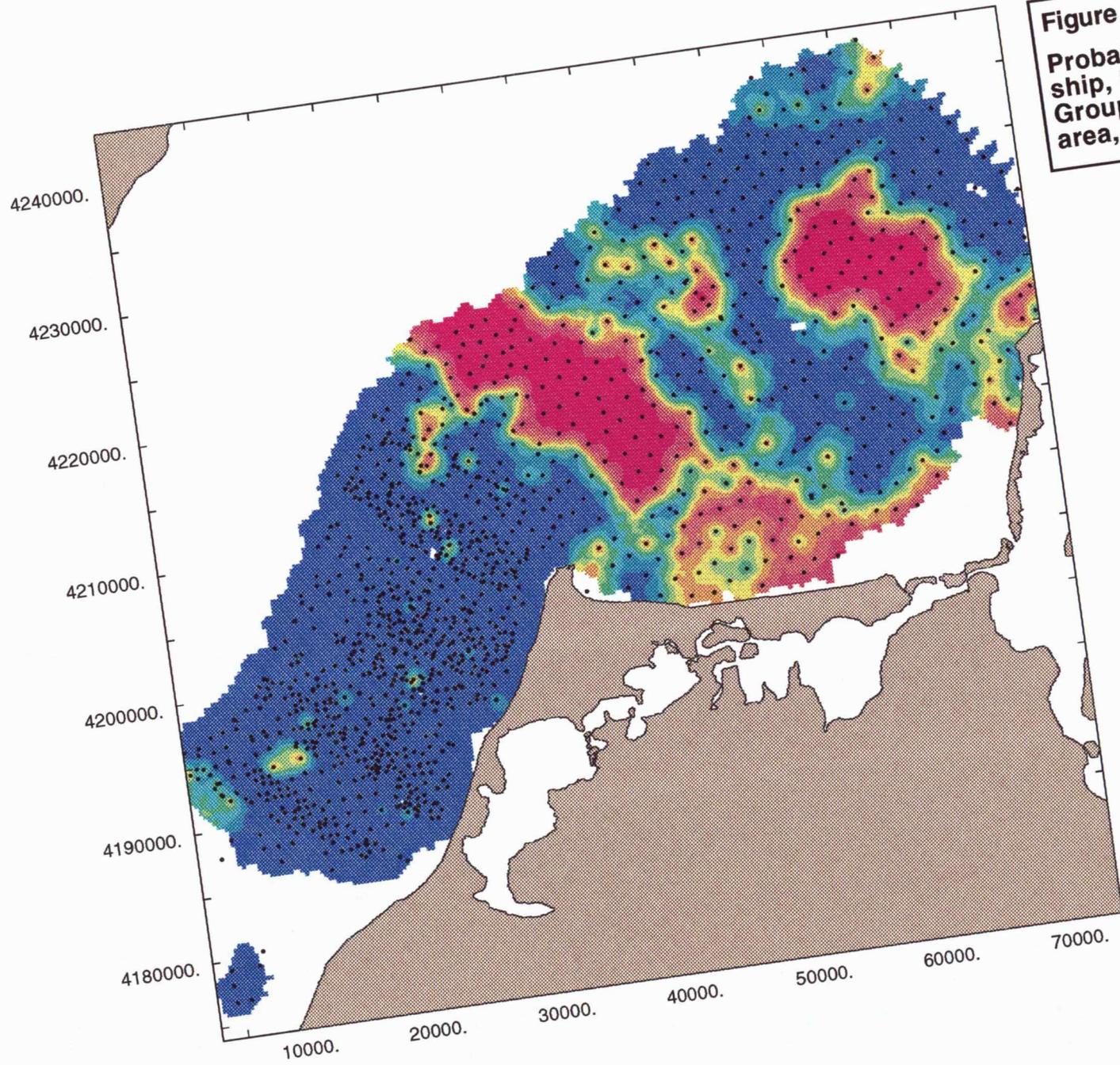
**Figure 11.**  
**Probability of member-**  
**ship, bottom sediment**  
**Group 3, Darss Sill**  
**area, Baltic Sea.**



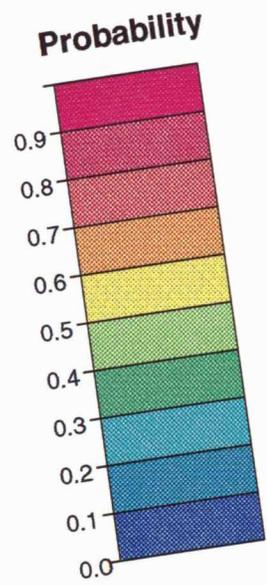


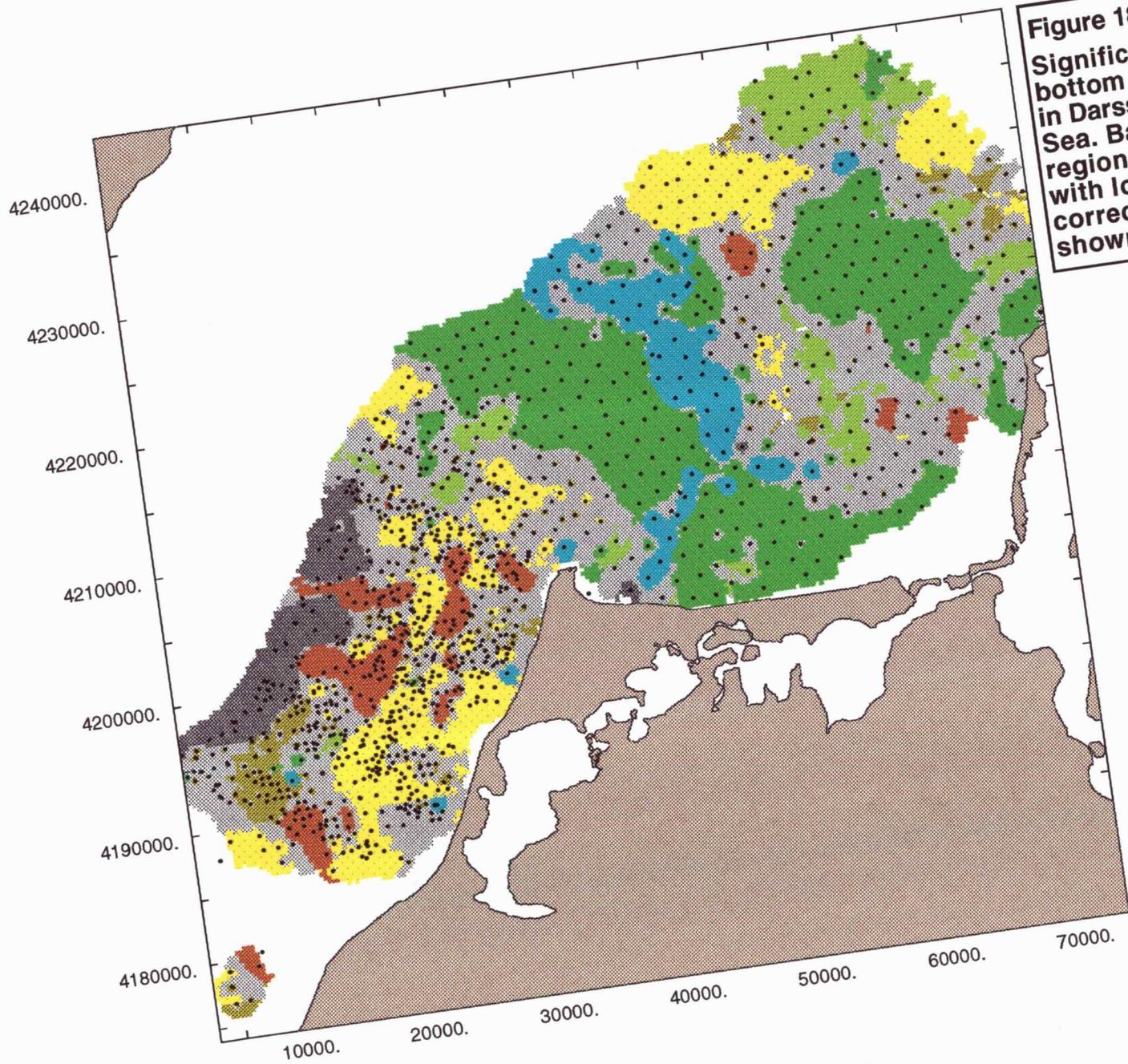
**Figure 12.**  
**Probability of membership,**  
**bottom sediment**  
**Group 4, Darss Sill**  
**area, Baltic Sea.**



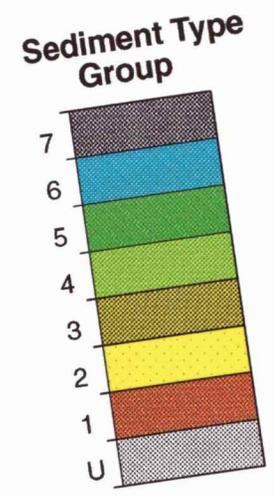


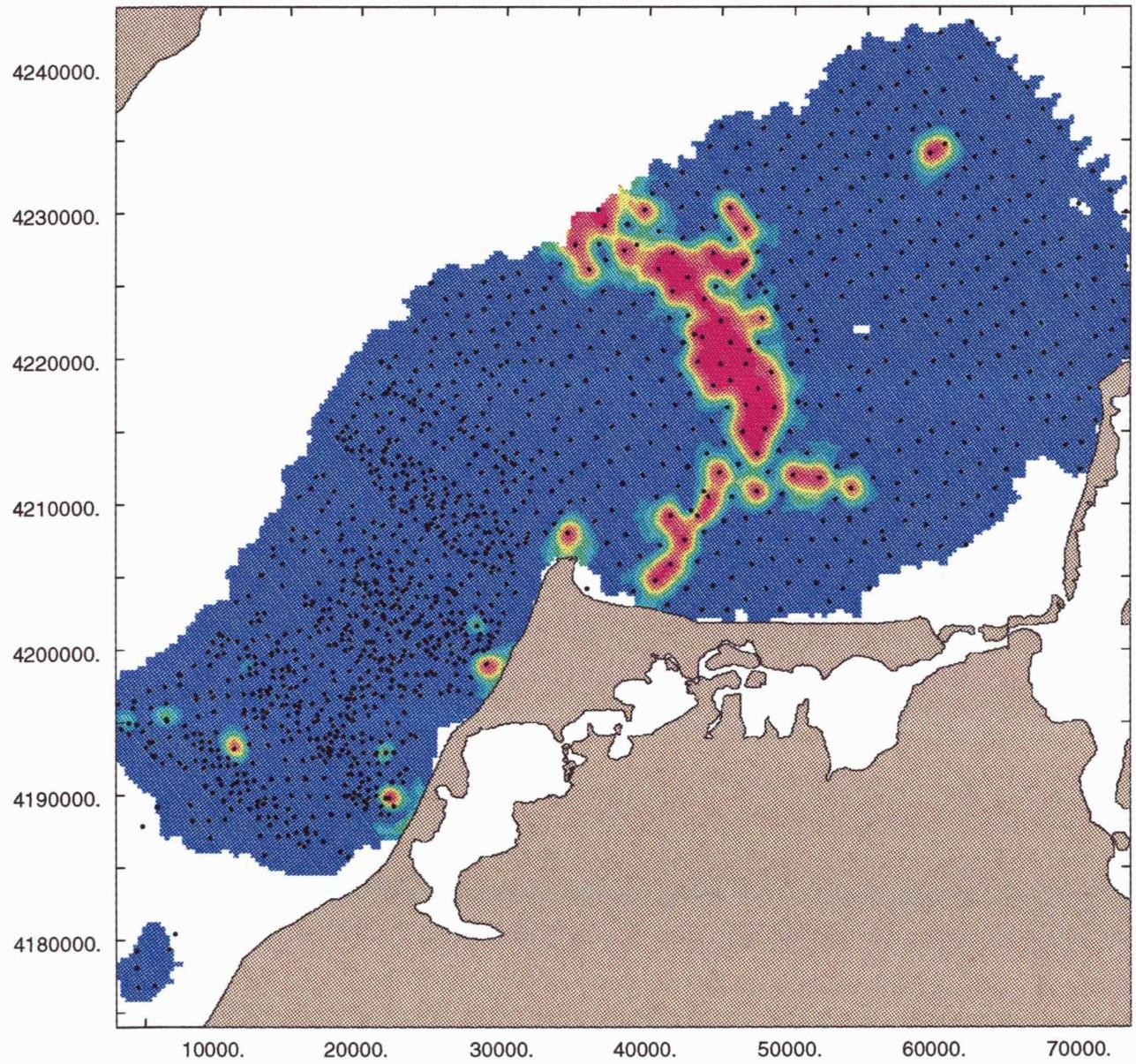
**Figure 13.**  
**Probability of membership,**  
**bottom sediment**  
**Group 5, Darss Sill**  
**area, Baltic Sea.**



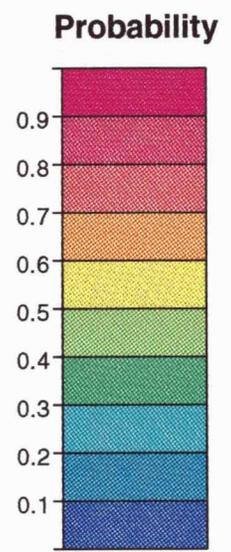


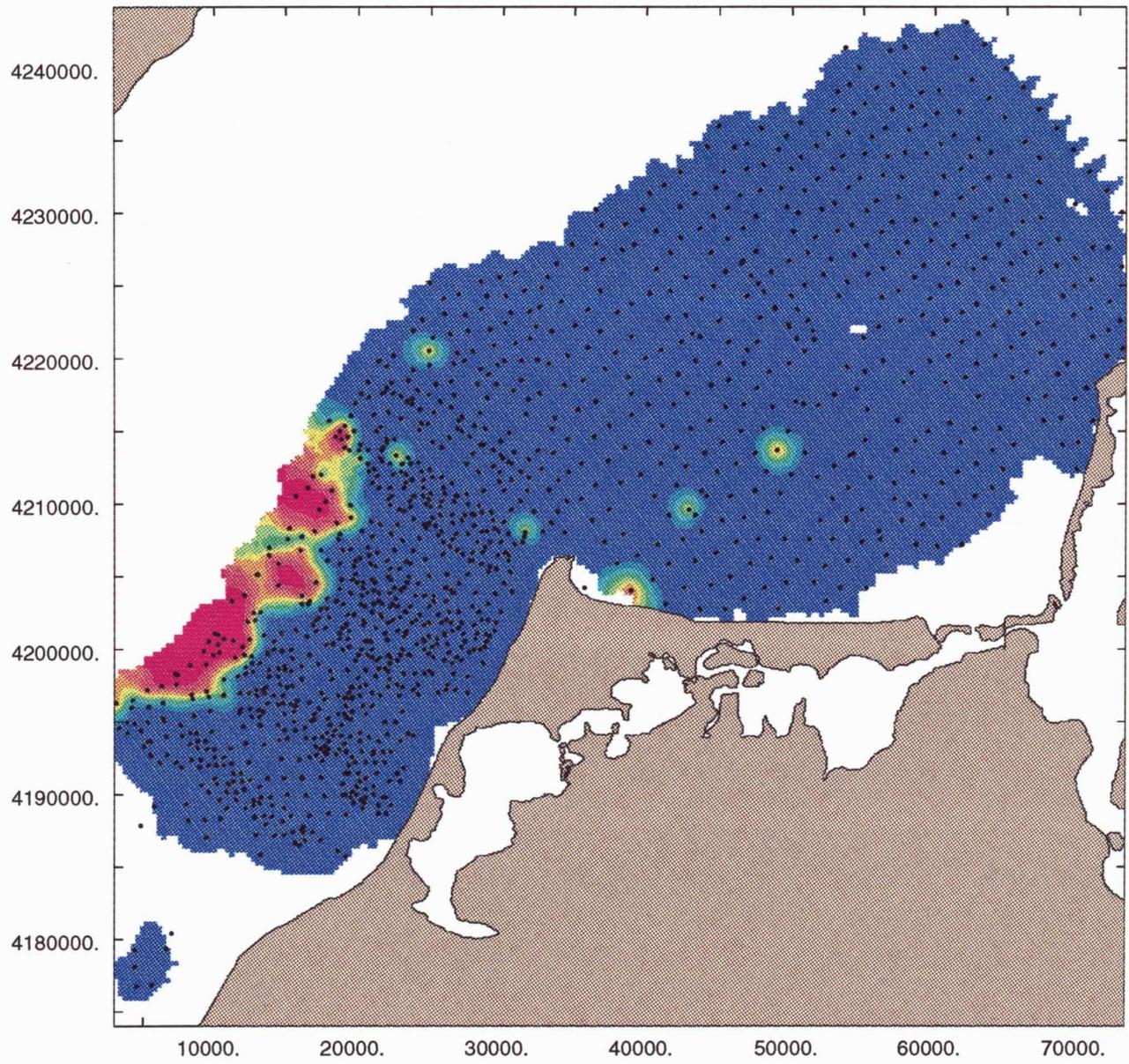
**Figure 18.**  
**Significant classes of bottom sediment types in Darss Sill area, Baltic Sea. Based on 7-class regionalization. Areas with low probability of correct classification shown in light gray.**



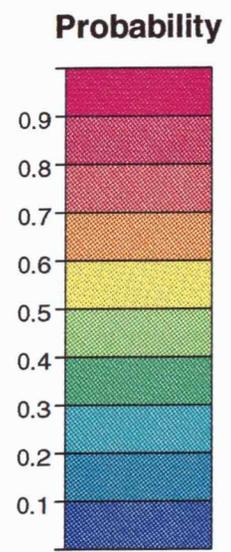


**Figure 14.**  
**Probability of membership, bottom sediment Group 6, Darss Sill area, Baltic Sea.**

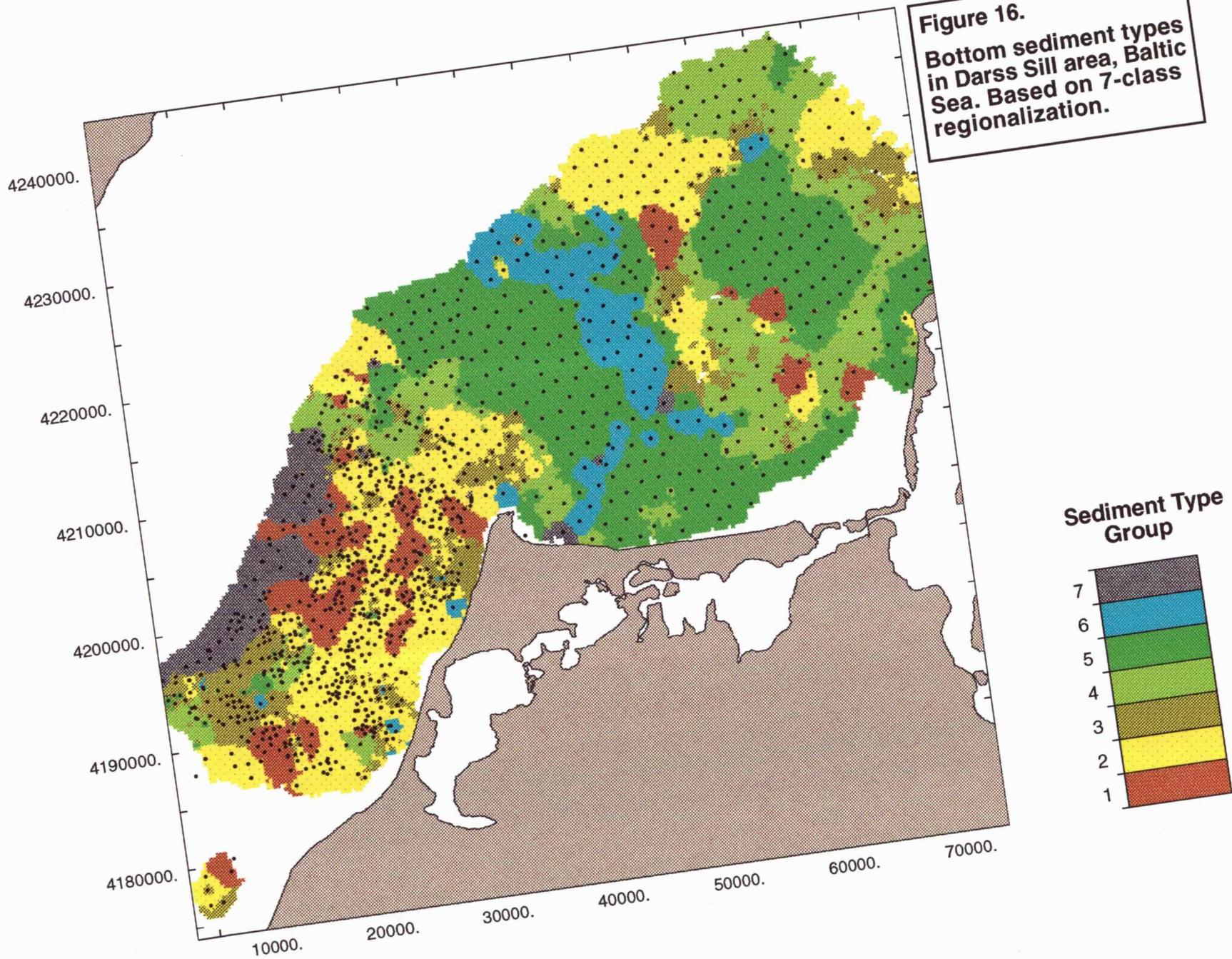


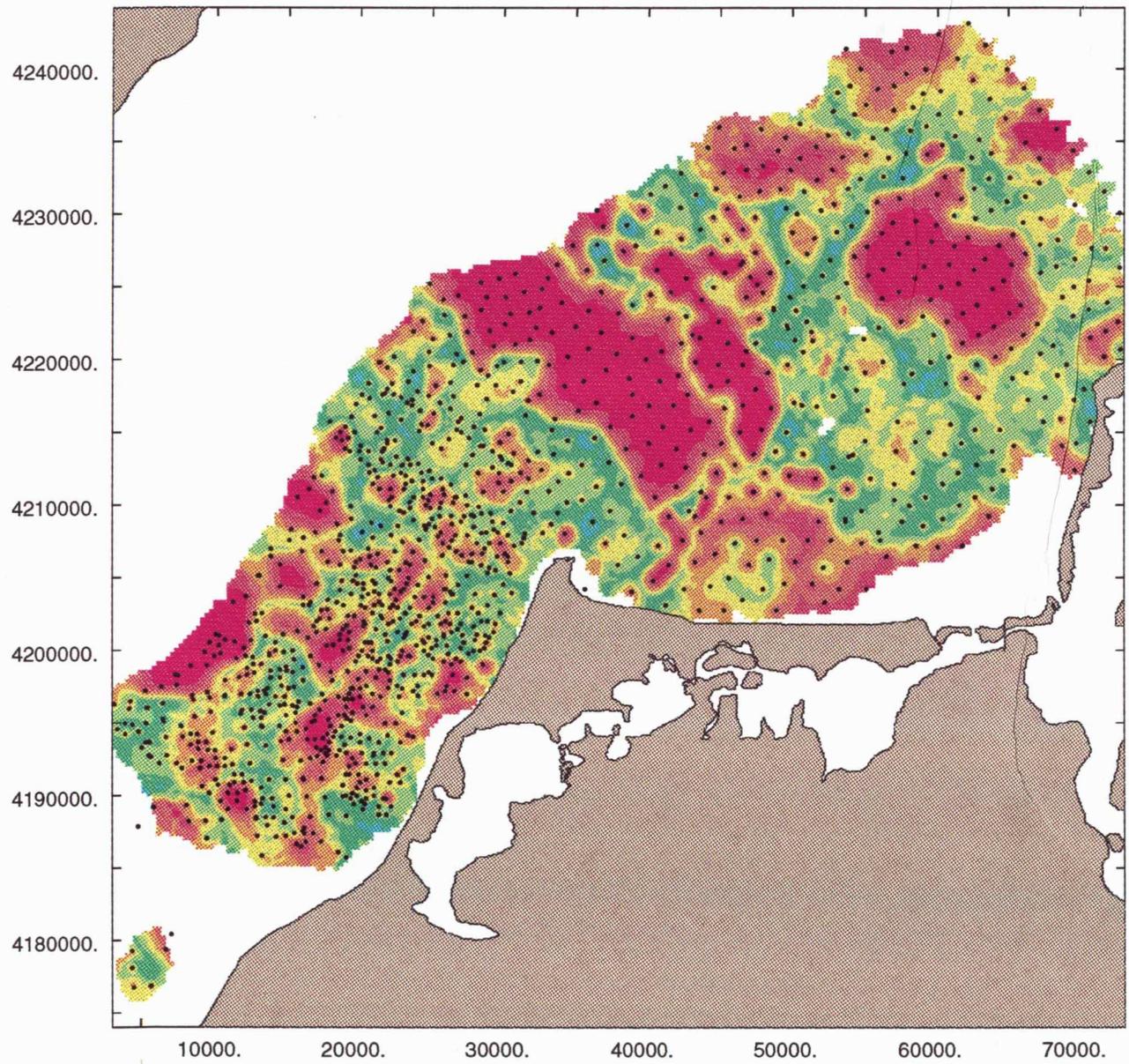


**Figure 15.**  
**Probability of membership, bottom sediment Group 7, Darss Sill area, Baltic Sea.**

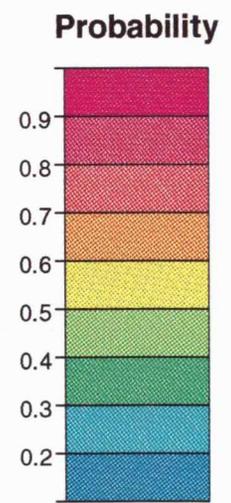


**Figure 16.**  
**Bottom sediment types**  
**in Darss Sill area, Baltic**  
**Sea. Based on 7-class**  
**regionalization.**





**Figure 17.**  
**Maximum probability of membership in any bottom sediment group, Darss Sill area, Baltic Sea.**



## COMPLICATIONS IN THE ANALYSIS

The Darss Sill sedimentological data have two characteristics that complicate their analysis. First, not all observations include all grain size intervals; in fact, only a small minority of the bottom samples contain grains in all size categories. This reflects the dynamics of sediment deposition; in areas of sufficient hydrodynamic energy to transport gravel-sized particles, fine sand and silt will be kept in suspension and swept away. Conversely, in areas of relatively still water where very fine sand and silt can settle out, there is insufficient energy to bring in large particles. Second, the Darss Sill sedimentological data are given in weight percent, and each observation sums to 100%. Such data are called “compositional data” and are said to be closed, because the quantity in one size category can only increase if there is a corresponding decrease in other categories. This constant sum constraint results in inherent negative correlations between some of the size categories; these are mathematical artifices of no physical significance.

These two characteristics have a similar mathematical effect. For some groups, there are certain grain size categories which are absent in all sediment samples. The covariance matrices of such groups will contain rows and columns whose elements are zero, and the determinants of the matrices will be zero. In other words, the covariance matrix of a sediment group with absent grain size categories will be singular.

In addition, the closure constraint imposes a collinearity of the form  $X_m = 100 - \sum_{i=1}^{m-1} X_i$  on the group covariance matrices. Although the negative dependencies are not apparent on casual examination because they are spread through all of the variables, they have the same effect as though there were a linear dependency between two rows (or two columns) of the matrix. A matrix with dependent rows or columns has a rank less than its size, is singular with a zero determinant, and has one or more zero eigenvalues.

Although we expect the covariance matrix of the Darss Sill data to be singular, initial calculations showed that this was not the case, although the covariance matrix was ill-conditioned. Upon examination, about 60 observations were noted that did not sum exactly to 100.0; most summed to 99.9 and a few summed to 100.1. These were made to sum to 100.0 by two alternative procedures, either by adjusting the weight percent of the most abundant size class up or down until the sum equaled 100.0, or by multiplying the weight percents in all size classes by  $(100/\text{sum})$ . The two adjustment procedures produced essentially identical results in subsequent analyses. Following adjustment, the covariance matrix of the Darss Sill data was singular.

Discriminant analysis is a variant of canonical analysis and is identical in computational form to multivariate analysis of variance (MANOVA). The coefficients of the discriminant functions are found from the eigenvectors of the product of the inverse of the pooled within-groups covariance matrix and the between-groups covariance matrix, that is, from  $\mathbf{E}^{-1}\mathbf{H}$ . This is analogous to finding the ratio of the between-groups sum of squares over the within-groups sum of squares in ordinary analysis of variance. The computational problem with the Darss Sill data (and all other collections of compositional data) arises because singular matrices such as the pooled within-groups covariance matrix cannot be inverted by conventional means.

The individual within-group covariance matrices each have an order of 8, representing the 8 grain-size classes of the sieve analyses. Of the seven individual within-group covariance matrices, the covariance matrix for Group 6 is singular and of rank 5, the covariance matrix for Group 1 is singular and of rank 6, and the covariance matrices of the other groups are singular and of rank 7. The pooled within-groups covariance matrix  $\mathbf{E}$  also is singular and of rank 7. The between-groups covariance matrix  $\mathbf{H}$  is singular and of rank 6, while the matrix product  $\mathbf{E}^{-1}\mathbf{H}$  is singular and of rank 7.

The discriminant functions are the eigenvectors associated with the matrix  $\mathbf{E}^{-1}\mathbf{H}$ . To determine these even in the face of the singularity of  $\mathbf{E}$ , we may turn to a singular value decomposition (SVD) algorithm. Here, the inverse of  $\mathbf{E}$  is found as the product of a square orthogonal matrix  $\mathbf{V}$  and its

*Analysis of Sedimentary Facies by Regionalized Classification*

transpose  $V^T$  and the inverse of a diagonal matrix  $W$ :

$$E^{-1} = VW^{-1}V^T$$

The inverse of a diagonal matrix is another diagonal matrix whose elements are reciprocals of the elements on the diagonal. The elements  $w_j$  are eigenvalues, and if any of these are zero or near zero, their reciprocal will be infinite or so large that the numerical solution to the equation becomes unstable. In the form of singular value decomposition used in the SAS numerical routines, these zero or near-zero elements are replaced by small constant values equal to a fraction of the average of the non-zero eigenvalues. The resulting solution is a *quasi-inverse* that will either satisfy or closely approximate the definition of an inverse:

$$E^{-1}E = I$$

Expressed in the form of SVD, the discriminant functions are found by solving the equation

$$E^{-1}H = VW^{-1}(V^TH)$$

The successive eigenvectors corresponding to the eigenvalues  $w_j$  are the columns of  $V$ . For the Darss Sill data, the numerical values of the matrices and their SVD solution are:

$E =$

8429.92325	2718.31218	1194.63848	-706.03981	-8058.4693	-2949.7983	-733.72594	105.159481
2718.31218	8743.49669	7219.40021	-4.7134598	-12400.955	-4967.87	-1268.4824	-39.1883
1194.63848	7219.40021	16434.8576	6909.93025	-17350.97	-10055.553	-2655.4043	-1696.899
-706.03981	-4.7134598	6909.93025	35012.5929	-13155.613	-20309.578	-5678.6186	-2067.9607
-8058.4693	-12400.955	-17350.97	-13155.613	136937.079	-36537.572	-26944.01	-22489.489
-2949.7983	-4967.87	-10055.553	-20309.578	-36537.572	100840.99	-17625.698	-8394.9207
-733.72594	-1268.4824	-2655.4043	-5678.6186	-26944.01	-17625.698	47646.6596	7259.27945
105.159481	-39.1883	-1696.899	-2067.9607	-22489.489	-8394.9207	7259.27945	27324.0191

$H =$

784.611983	2047.22829	4074.062	6950.76886	2017.72839	-17735.452	-620.15168	2481.20433
2047.22829	6014.92295	12183.818	21386.3416	9221.93579	-46664.856	-4996.0058	806.615592
4074.062	12183.818	24822.7637	44165.5378	26516.4881	-99331.803	-11884.395	-546.47162
6950.76886	21386.3416	44165.5378	81247.7909	85599.3265	-203035.43	-28451.301	-7863.0345
2017.72839	9221.93579	26516.4881	85599.3265	661630.505	-608441.52	-117244.87	-59299.598
-17735.452	-46664.856	-99331.803	-203035.43	-608441.52	940887.235	62968.8731	-28647.046
-620.15168	-4996.0058	-11884.395	-28451.301	-117244.87	62968.8731	65808.1625	34419.6829
2481.20433	806.615592	-546.47162	-7863.0345	-59299.598	-28647.046	34419.6829	58648.6468

$W =$

10.9345867	0	0	0	0	0	0	0
0	2.38636748	0	0	0	0	0	0
0	0	1.89191117	0	0	0	0	0
0	0	0	0.71575369	0	0	0	0
0	0	0	0	0.00219194	0	0	0
0	0	0	0	0	0.00026081	0	0
0	0	0	0	0	0	0.000000002	0
0	0	0	0	0	0	0	0

The matrix  $E$  is not exactly singular as theoretically expected, presumably because of the accumulation of rounding errors during calculation of the sums of squares and cross products. The matrix

is, however, poorly conditioned and its inversion requires that either a Moore-Penrose generalized inverse or a quasi-inverse be calculated. In this instance, either procedure will yield results that differ only by a constant.

The eigenvectors associated with eigenvalues of  $\mathbf{E}^{-1}\mathbf{H}$  that are identically equal to or very near zero can be discarded without effecting the remaining eigenvectors; when normalized, these constitute the desired linear discriminant functions (McLachlan, 1992). The contribution of an individual discriminant function to the separation between the groups is expressed by the proportion of the trace contributed by its corresponding eigenvalue. The eigenvalues are given as the diagonal elements of  $\mathbf{W}$ ; those marked with an asterisk are deleted from further calculations.

$w_j$	Percent
10.9346	68.6%
2.3864	15.0%
1.8919	11.9%
0.7158	4.5%
0.0167	0.0%
0.0022	0.0%
0.0003*	0.0%
0*	0.0%

The normalized eigenvectors can be used to project the observations and the sediment group means onto the canonical axes. Figure 19 is a projection onto the first two axes. As expected, the canonical projections of the group centroids are completely distinct, especially along the first axis.

For regionalization, it is necessary to express the difference between each observation and the centroids of the groups in terms of Mahalanobis' distance,  $D^2$ . In turn, these distances are converted into posterior probabilities of membership in each of the groups. Let  $\mathbf{y}_i$  be the vector of weight percents for sediment sample  $i$ . Mahalanobis' distance between this observation and the centroid of group  $k$ , represented by the vector  $\mathbf{M}_k$ , is:

$$d_{i,k} = \mathbf{y}_i' \mathbf{E}^{-1} \mathbf{y}_i - 2\mathbf{y}_i' \mathbf{E}^{-1} \mathbf{M}_k + \mathbf{M}_k' \mathbf{E}^{-1} \mathbf{M}_k$$

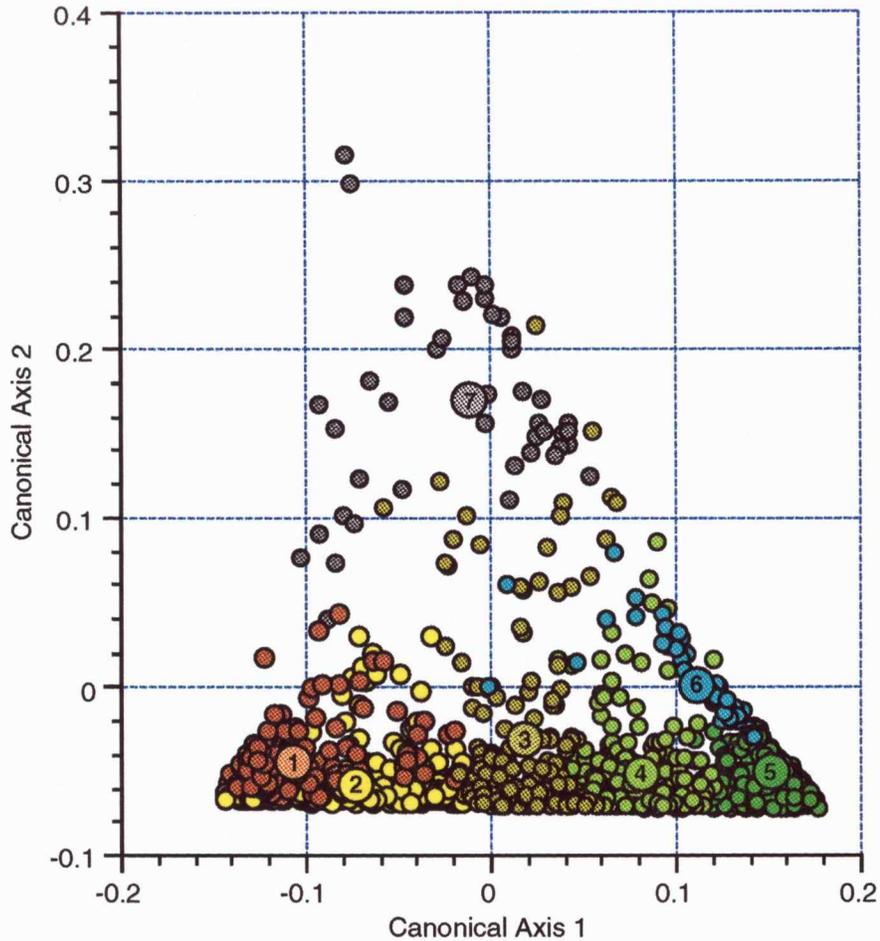
If we assume each group has a multivariate normal distribution, the posterior probability (Gnedenko, 1970) that an observation belongs to group  $k$  is:

$$p(k | i) = \frac{\exp(-0.5d_{i,k})}{\sum_1^k \exp(-0.5d_{i,k})}$$

These are the probabilities that are mapped to define areas of membership in the sediment groups of the Darss Sill region.

The Darss Sill sedimentological data obviously are not multivariate normal, one of the presumptions underlying all multivariate analysis procedures, including discriminant functions. In addition to the skew introduced by the absence of some grain size categories, compositional data cannot be normally distributed because they are constrained to a range. However, multivariate normality is not required for estimation and data description, although it is essential for inference (Marascuilo and Levin, 1983). In addition, linear discriminant analysis assumes the individual within-group covariance matrices are equal. Tests of homogeneity confirm what can be anticipated from the differences in rank of the within-group covariance matrices—they are not equal. Because of inhomogeneity, the computed linear discriminant functions cannot be optimal, but the degree and practical significance of non-optimality cannot be assessed. The discriminant functions could be calculated in a quadratic form that accommodates unequal individual within-group covariance matrices, but tests show that

*Analysis of Sedimentary Facies by Regionalized Classification*



**Figure 19.** Projection of bottom sediment observations from the Darss Sill area of the Baltic Sea onto their first two canonical axes. Circled numbers indicate group centroids.

for the Darss Sill data, the misclassification rate of a quadratic discriminator is significantly higher than the rate for a linear discriminator. Undoubtedly the Mahalanobis' distances between observations and group centroids are incorrect as a consequence of erroneously assuming homogeneity, and the posterior probabilities of group classification are therefore incorrect as well. However, for most of the observations the calculated posterior probabilities of group membership are extreme (near 1.0 or 0.0), reflecting the distinctness of the sediment groups along the canonical axes. In view of this, the potential effect of distortions in the posterior probabilities seems inconsequential.

These complications could be avoided if it were possible to transform the Darss Sill data to normal form. Most compositional data can be normalized using Aitchison's (1986) centered logratio transformation, but this is not possible with the Darss Sill data. There are significant numbers of observations in all groups that do not contain grains in one or more size classes. For example, none of the observations in Groups 1 and 6 contain gravel-sized particles and, in addition, none of the observations in Group 6 contain coarse sand. Many of the observations in Group 7 contain no silt-sized material. Such zero entries cannot be transformed by an operation involving logarithms. To apply the logratio transformation, it would be necessary to either combine grain sizes into no more than four classes, or to work with a subcomposition of only four classes. Since both the compacted data and the subcomposition also are subject to closure, only three degrees of freedom would be

available for distinguishing groups. The most distinctive aspects of the sedimentological data would have been discarded.

In an unpublished report, Zhou Di and others applied Aitchison's centered logratio transformation to the Darss Sill data by replacing zero values in the data with an arbitrary small constant, thus avoiding the problem of taking logarithms of zero. Unfortunately, the results seem highly sensitive to the choice of constant.

## SUMMARY

The method of regionalized classification developed by Harff and Davis (1990) can be used to subdivide clastic sedimentary facies on the basis of grain size measurements. The procedure is carried out in two steps. First, the sediments are grouped based on similarities in the grain size distributions of samples using a method of hierarchical unsupervised classification. The groups are described by their average grain size distributions. A genetic interpretation of the groups is based on a cross-plot of average grain size and sorting. The succession of groups can be interpreted as a genetic sequence extending from source areas to depositional sites. In a second step, the regional distributions of the groups are expressed by a spatial function of the probability of group membership. The probabilities are regarded as stochastic regionalized variables and kriging interpolation algorithms are used to construct a probability map for each group. Each probability map shows patterns in the distribution of a sediment type, permitting a genetic interpretation related to a specific depositional environment. By considering these maps simultaneously, any location can be assigned to membership in the group to which it has the highest probability of occurrence. Because the allocation is performed on a regular grid, it is easy to create a geological map of the distribution of facies types in an area of investigation using computer mapping algorithms.

Regionalization of bottom sediment data from the Darss Sill area of the Baltic Sea—the bottleneck in the connection of this marginal sea with the ocean—clearly delineates source areas that provide poorly sorted, coarse grained sediments eroded from tills that crop out on the sea floor, the routes of transport of these sediments, and the depocenters where they come to rest. During transport, the sediments are progressively separated and eventually become well-sorted, fine-grained facies. The succession is represented by a genetic chain from Group 1 sediments to Groups 5 and 6. The intermediate sediment Groups 2, 3 and 4 are transition types. The western part of the Darss Sill area is dominated by source and lag sediments, and depocenters dominate in the eastern part, reflecting the direction of movement of the main currents. The easternmost depocenter for Group 5 shows a circular pattern surrounded by transitional groups, suggesting that a pattern of circular currents (gyres) were responsible for sediment transport. Sediment Group 7 is a poorly sorted, heterogeneous group reflecting the relatively unsettled conditions within the Kadet Channel.

The success of regionalization in producing interpretable facies patterns in the Darss Sill area of the Baltic Sea indicates that the methodology may be useful in many similar circumstances, involving ancient sediments as well as modern. Although many of the distributional assumptions underlying regionalization are violated by the use of compositional data such as grain size measurements, regionalization seems to be robust and produces acceptable results. It is noteworthy that discrimination and calculation of Mahalanobis' distances can be performed using singular value decomposition even though use of compositional data inevitably means that the covariance matrices are singular. Although the calculated probabilities may be distorted, the results of this study of the Darss Sill suggests that these theoretical problems may have limited practical effect. However, both the theoretical problems and our results should encourage others to investigate alternatives.

*Analysis of Sedimentary Facies by Regionalized Classification*

REFERENCES

- Aitchison, J., 1986, *The Statistical Analysis of Compositional Data*: Chapman and Hall, London, 416 p.
- Bock, H.H., 1974, *Automatische Klassifikation*: Vandenhoeck & Rupprecht, Gottingen, 480 p.
- Davis, J.C., 1986, *Statistics and Data Analysis in Geology*: John Wiley & Sons, New York, 646 p.
- Deutsch, C.V., and Journel, A.G., 1992, *GSLIB—Geostatistical Software Library and User's Guide*: Oxford Univ. Press, New York, 340 p.
- Fisher, R.A., 1936, The use of multiple measurements in taxonomic problems: *Annals of Eugenics*, no. 67, p. 179–188.
- Gnedenko, B.W., 1970, *Lehrbuch der Wahrscheinlichkeitsrechnung*: Akademie Verlag, Berlin, 399 p.
- Griffiths, J.C., 1967, *Scientific Method and the Analysis of Sediments*: MacGraw-Hill, Inc., New York, 508 p.
- Harff, J., and Davis, J.C., 1990, Regionalization in geology by multivariate classification: *Mathematical Geology*, vol. 22, no. 5, p. 573–588.
- Harff, J.E., Davis, J.C., and Eiserbeck, W., 1993, Prediction of hydrocarbons in sedimentary basins: *Mathematical Geology*, v. 25, no. 7, p. 925–936.
- Journel, A.G., and Huijbregts, C.J., 1978, *Mining Geostatistics*: Academic Press, London, 600 p.
- Lemke, W., 1992, Zur Korngrößenverteilung psammitischer Sedimente im Bereich der Darßer Schwelle: *Zeitschrift für geologische Wissenschaften*, Band 20 (5/6), S. 483–492.
- Lemke, W., Kuijpers, A., Hoffmann, G., Milkert, D., and Atzler, R., 1994, The Darß Sill, hydrographic threshold in the southwestern Baltic, Late Quaternary geology and recent sediment dynamics: *Continental Shelf Research*, v. 14, no. 7/8, p. 847–870.
- Marascuilo, L.A., and Levin, J.R., 1983, *Multivariate Statistics in the Social Sciences—A Researcher's Guide*: Brooks/Cole Publishing Co., Monterey, Calif., 530 p.
- McLachlan, G.J., 1992, *Discriminant Analysis and Statistical Pattern Recognition*: John Wiley & Sons, New York, 526 p.
- Rodionov, D.A., 1981, *Statisticěskie resěnija v geologii.*: —Izd. "Nedra", Moskva, 231 p.
- Tatsuoka, M.M., 1971, *Multivariate Analysis: Techniques for Educational and Psychological Research*: John Wiley & Sons Inc., New York, 310 p.
- Voronin, Ju. A., 1967, *Geologija i matematika*: —Izd. "Nauka", Novosibirsk, 253 p.