Regionalization of Western Kansas Based on Multivariate Classification of Stratigraphic Data from Oil Wells, II

J. Harff,[1] J. C. Davis,[2] R. Olea[2]
and G. Bohling,[2]

Kansas Geological Survey
Open File Report #91-40

[1]Zentralinstitut für Physik der Erde, Potsdam 1561, BRD.
[2]Kansas Geological Survey, The University of Kansas, Lawrence, KS 66047 USA.

CONTENTS

TABLE

ILLUSTRATIONS

ILLUSTRATIONS

Regionalization of Western Kansas Based on Multivariate Classification of Stratigraphic Data from Oil Wells, II

J. Harff,[1] J. C. Davis,[2] R. Olea[2]
and G. Bohling,[2]

Abstract

Regionalized classification in geology requires interpolating between observation points such as drill locations. Rather than kriging the *n*-dimensional variable of geological properties, it is preferable either to krige the probabilities of correct classification estimated at the observation points or to krige the Mahalanobis' distances between observations and class centroids. The second method is called **normalized distance kriging**. The probability of correct classification can easily be calculated by assuming the kriging error distribution of the kriged normalized distances is Gaussian. This method is advantageous because it incorporates the uncertainty of spatial prediction caused by irregular distributions of observations, and it requires relatively simple calculations compared to probability kriging. A demonstration using data from the Western Kansas Shelf illustrates the method's utility for prediction of undiscovered oil resources.

Introduction

The method of regionalized classification was developed by Harff and Davis (1990) and was applied to a regional study of the western Kansas shelf, as documented in the first open-file report on this topic (Harff and others, 1989). Although these papers examined the general concept of

[1]Zentralinstitut für Physik der Erde, Potsdam 1561, BRD. Present Address: c/o Kansas Geological Survey, The University of Kansas, Lawrence, KS 66047 USA.

[2]Kansas Geological Survey, The University of Kansas, Lawrence, KS 66047 USA.

regionalization, the theoretical basis of the important problem of interpolation was not addressed.

A joint research program between the Mathematical Geology Sections of the Kansas Geological Survey (KGS) and the Central Institute for Physics of the Earth in Potsdam (ZIPE) was begun under the direction of Dr. J. Harff and Dr. J.C. Davis in 1990, to investigate the application of the regionalization concept. The intent is to confirm the concept theoretically and to apply the methodology for modeling in the Kansas intracontinental basin and the northeast German basin. With sponsorship from the German Science Foundation, J. Harff of ZIPE served as Visiting Research Scientist at the KGS from March through September of 1991, and members of the ZIPE staff worked in Lawrence during September and October, 1991. Initial results of the application of regionalization on a regional scale are described in Watney and others (in press) and Harff, Davis and Olea (1991). This report focuses on the interpolation and decision-making problems of regionalization. Results from the three-dimensional regionalization of the Zenith oil field in western Kansas are documented separately (Harff and others, 1991).

Regionalization

Model of Regionalization

The regionalization concept presupposes that a partition of the Earth's crust can be divided into "homogeneous" subunits separated by stratigraphic boundaries, facies boundaries, tectonic faults, or other discontinuities. In general terms, a geologic body B is a part of the Earth's crust. The features of this body, projected onto the Earth's surface or a subsurface datum, can be modeled as a random function

$$X(r) = m(r) + Y(r), \quad \forall r \in R, \tag{1}$$

where $x(r) \in X(r)$ denotes an n-dimensional random variable describing the features of the geological body, $m(r)$ denotes its deterministic parts, $Y(r)$ represents its stochastic components, and $r \in R$ describes the projection area. The covariance matrix is

$$\Sigma(r) = \mathrm{E}\left[(X(r) - m(r))(X(r) - m(r))'\right], \quad \forall r \in R. \tag{2}$$

The body B is subdivided into homogeneous bodies $B_{ij} \in B$, $i \in I, I = \{1,\ldots,K\}, j \in J_i, J_i = \{1,\ldots,N_i\}$. R_{ij} is the subregion of the projection area R that corresponds to part B_{ij}. Within these homogeneous parts, the expected value function m is assumed to be constant, as is the covariance matrix. That means for each homogeneous body, $m(r) = m_i$ and $\Sigma(r) = \Sigma_i, \forall r \in R_{ij}$, where $m_i \in M$, $M = \{m_1,\ldots m_K\}$ and $\Sigma_i \in \breve{\Sigma}$, $\breve{\Sigma} = \{\Sigma_1,\ldots \Sigma_K\}$. This implies that the random field $X(r)$ is stationary within each subregion R_{ij}.

The projection of the boundary of the homogeneous bodies onto R is described as

$$L\left(B_{ij}\right) = \left\{ r \in R_{ij} : \exists\, r' \in R : r \overset{N}{\sim} r' \wedge m(r) \neq m(r') \right\} \quad \forall i \in I, \forall j \in J_i \tag{3}$$

where $r \overset{N}{\sim} r'$ indicates a general neighborhood relation (see Haslett, 1989). Pursuing the discrete concept of regionalization, a homogeneous body B_{ij} can be subdivided into a finite number of elementary bodies $b_k \in B_{ij}$, $k \in \left\{1,\ldots,N_{ij}\right\}$ each of which is characterized by the same distribution law of the n-dimensional variable X. These elementary bodies are the smallest units to be studied. The size chosen for these units depends upon the storage capacity and execution speed of the hardware available for numerical solutions.

The boundaries of the elementary units are usually determined as edges of cells in a regular grid. Therefore, for the geographic determination of the elementary bodies, it is sufficient to refer to the centers of these cells, which in general we can take as the nodes $r \in R$ of a grid. In this conceptualization, a geological body B can be regarded as consisting of a hierarchically structured set of classes B_{ij} of elementary geological bodies:

$$B = \bigcup_{i \in I} \bigcup_{j \in J_i} B_{ij} \tag{4}$$

These classes can be merged on the basis of their stochastic parameters

$$B_i = \bigcup_{j \in J_i} B_{ij}, \quad \forall i \in I \tag{5}$$

$$B_i = \{ b \in B : m(b) = m_i \wedge \Sigma(b) = \Sigma_i \}, \quad m_i \in M, \ \Sigma_i \in \breve{\Sigma}. \tag{6}$$

The set of classes B_i forms a partition Z

$$Z = \{ B_1, B_2, \ldots, B_K \}. \tag{7}$$

Estimation at the Unsampled Locations

In typical geological situations, experimental data are available only for a relatively small sample $b \in B' \subset B$. These data come from well records, observations of outcropping rocks, or geophysical measurements. The data are stored as ordered matrices

$$\begin{pmatrix} x_1(b(r_1)) & \ldots & x_n(b(r_1)) \\ \cdot & \ldots & \cdot \\ \cdot & \ldots & \cdot \\ \cdot & \ldots & \cdot \\ x_1(b(r_N)) & \ldots & x_n(b(r_N)) \end{pmatrix}. \tag{8}$$

The partition Z for the set B' may be given, or may be determined by an unsupervised classification procedure. Estimates m_i^* and S_i of the expected value vector and the covariance matrix can be determined for each of the classes B_i using standard statistical methods.

The regionalization of an area also requires the allocation of the elementary bodies $b \in \overline{B}'$ which belong to the complementary class of the random sample, $\overline{B}' = B \setminus B'$. After classification of the complete set B, homogeneous regions R_{ij} can be determined by searching for geological bodies belonging to the same class B_i which are also neighbors. However, for the complementary set $\overline{B}'$ no measurement data are available, so the regionalization requires, as do most geological tasks, a step of interpolation or spatial prediction.

To interpolate an n-dimensional nonstationary random field, the cokriging approach of Myers (1982) is available. However, because of the numerical complications of this approach (Harff and Davis, 1990; Harff, Davis and Olea, 1991), we propose instead to transform the n-dimensional variable space into a "decision space" of lower dimension and to interpolate "discriminators" derived by this transformation.

Bayesian Transformation

Given is a set B' of elementary geologic bodies and a partition of B',

$$Z' = \{B_1, B_2, \ldots, B_k\}.$$

Each class consists of elements $b_{ij} \in B_i, \forall i \in I, \ \forall j \in O_i, \ O_i = \{1, \ldots, F_i\}$, for which measurement values $\left(x(b_{ij})\right)$ are available. These data represent a mixture $P(x) = \sum_{i \in I} p_i \cdot p[x|i]$ of K n-dimensional probability distributions $p[x|i]$. The *a priori* probability of class i is designated as p_i. From the data, the following statistics can be estimated: The sets of experimental means $M^* = \{m_i^*\}$ and experimental intraclass covariance matrices $S=\{S_i\}, \forall i \in I$; the matrix G of unstandardized discriminant coefficients; and the vector $\boldsymbol{a}$ of associated constants (Fisher, 1936; Tatsuoka, 1971). An n-dimensional variable $f \in F$, the canonical discriminant function with value vectors $f(b_{ij})$, can be derived by transformation of X

$$f(b_{ij}) = x(b_{ij})\boldsymbol{G} + \boldsymbol{a}, \; \forall i \in I, \; j \in O_i. \tag{9}$$

For these variables, experimental mean vectors $C^* = \{c_i^*\}$ and experimental intraclass covariance matrices $D=\{D_i\}, \forall i \in I$ can be estimated.

Using Bayes' formula, the probability of class membership for each of the elementary geological bodies $\forall b \in B'$ can be determined using either the original variable X or the transformed variable F. For the original variable the equation has the following form:

$$p[i|x] = \frac{p_i |S_i|^{-1/2} \exp\left(-d_i^2 / 2\right)}{\sum_{j \in I} p_j |S_j|^{-1/2} \exp\left(-d_j^2 / 2\right)}, \tag{10}$$

in which d_i^2 is Mahalanobis' distance

$$d_i^2 = \left(x(b) - m_i^*\right)' S_i^{-1} \left(x(b) - m_i^*\right). \tag{11}$$

A complementary equation uses the canonical discriminant score

$$p[i|f] = \frac{p_i |D_i|^{-1/2} \exp\left(-\chi_i^2 / 2\right)}{\sum_{j \in I} p_j |D_j|^{-1/2} \exp\left(-\chi_j^2 / 2\right)} \tag{12}$$

with

$$\chi_i^2 = \left(f(b) - c_i^*\right)' D_i^{-1} \left(f(b) - c_i^*\right). \tag{13}$$

The transformation procedure for variable X forms new variables d_i^2, χ_i^2, $p[i|x]$, and $p[i|f]$, and implies a reduction of the dimensionality of the variables from n to 1.

Interpolation and Decision Making

For an interpolation procedure whose objective is decision-making at locations where no measurement data are available, the variables obtained by transformation of the original variable X can be regarded as regionalized variables. This assumption is justified if the original variable X is assumed to be a regionalized variable. Let the symbol ρ stand for one of the variables d_i^2, χ_i^2, $p[i|x]$ or $p[i|f]$, $\forall i \in I$.

As in the case of $X(r)$,

$$\rho(r) = m'(r) + Y'(r), \forall r \in R \tag{14}$$

is assumed to be nonstationary, but to exhibit stationarity in homogeneous subregions R_{ij}. It is assumed that the drift may be modeled locally by a polynomial

$$m'(r) = \sum_{l=0}^{k} a_l \, g^l(r), \forall r \in R. \tag{15}$$

$Y'(r)$ denotes the stochastic component whose spatial correlation structure is described by the semivariogram

$$\gamma(h) = \tfrac{1}{2} \mathrm{E}\left[\left(Y'(r+h) - Y'(r)\right)^2\right]. \tag{16}$$

If observations $x\left(b(r_j)\right)$ exist at a limited number of locations $j \in J$, $\rho(r_j)$ can be assessed and an estimator may be determined at a location r_e where no observation is available

$$\rho^*(r_e) = \sum_{j \in J} \lambda_j \cdot \rho(r_j). \tag{17}$$

λ_j is determined using a universal kriging system of equations

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B} \tag{18}$$

with

$$\mathbf{X} = (\ldots\lambda_i\ldots, \quad \ldots\mu_i\ldots)' , \tag{19}$$

$$\mathbf{B} = \left(\ldots\gamma(r_i, r_e)\ldots, \quad \ldots g^i(r_j)\ldots\right)' , \tag{20}$$

and

$$\mathbf{A} = \begin{pmatrix} \vdots & \vdots & \vdots \\ \ldots\gamma(r_i, r_j)\ldots & \ldots 1 \ldots & g^i(r_j)\ldots \\ \vdots & \vdots & \vdots \\ \ldots 1 \ldots & 0 & \\ \vdots & \vdots & 0 \\ g^i(r_j) & 0 & \\ \vdots & \vdots & \end{pmatrix} . \tag{21}$$

μ_i in vector $\mathbf{X}$ is a Lagrangian multiplier. The kriging variance is determined by

$$\sigma^2(r_e) = \mathbf{X}'\mathbf{B} . \tag{22}$$

The probability that an elementary body $b(r_e)$ can be allocated to a class can be determined either from the kriged probabilities ($p[i|x(r)]$ or $p[i|f(r)]$), or by the kriged distances ($d_i^2(r)$ or $\chi_i^2(r)$), and calculating the Bayes' probabilities based on the kriged distances.

The body $b(r_e)$ will be allocated to the class B_i for which the class membership probability of $b(r_e)$ is a maximum. An alternative approach is to krige the Mahalanobis' distances and allocate the body $b(r_e)$ to a class B_i using the minimum distance principle (Tatsuoka, 1971), taking into account the kriging variance. Both possibilities are discussed here, by an example in which the variables $p[i|f(r)]$ and $d_i^2(r)$ calculated for a data set from western Kansas.

Data for the Case Study

The data set consists of the records from 1245 exploratory holes drilled in western Kansas. The well localities are classified to form a regionalization of the western Kansas shelf area, and the probabilities of the discovery of oil are estimated based on the success ratios within the regions. The stratigraphic interval of investigation includes the Lansing and Kansas City Groups of Pennsylvanian (Upper Carboniferous) age, which consist of interbedded marine limestones and shales that were deposited in a shallow epicontinental sea that covered the North American midcontinent in late Paleozoic time. Oil has been produced from limestone reservoirs in these rocks for many decades; the local petroleum industry is in a mature stage of development. Thousands of wells have been drilled in the search for oil, and hundreds of fields have been discovered. A general description of the petroleum geology of the study area is given by Watney (1984) and Watney and others (in press). Figure 1 is an index map showing the locations of oil fields in Kansas and the extent of the study area.

Of the 1245 exploratory drill holes that penetrate the Lansing–Kansas City, 917 were dry, 99 discovered oil in the Lansing–Kansas City interval, and 229 were "unclassified," with unknown producing status. Well locations are posted in Figure 2. Geological information consists of the subsea elevations of the top of key formations, the thicknesses of selected stratigraphic intervals, and petrophysical properties measured on well logs and averaged over selected intervals. In this example, the mutually exclusive classes dry or oil producing were used. Z was determined as $\{B_{\text{oil}}, B_{\text{dry}}\}$, $b_k \in B_{\text{oil}}$, $k \in \{1,\ldots,497\}$, $b_k \in B_{\text{dry}}$, $k \in \{1,\ldots,748\}$.

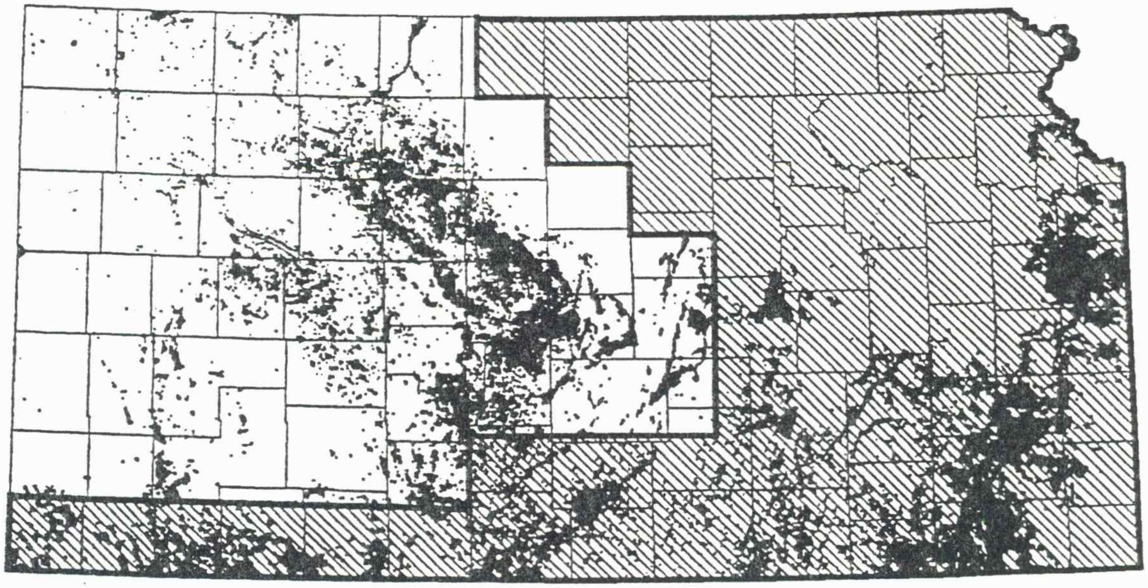

Figure 1. Oil fields in Kansas. Unruled area indicates the part of Kansas covered by this study.

The original data set contains sixteen geological variables; preliminary studies indicated that some are either redundant or only weakly related to oil production. After calculating a linear discriminant function between the two groups, Fisher's criterion was used to select the effective discriminating variables; six proved to have significant discriminating power and are listed below. Terms such as "K-zone" or "I-zone" refer to specific limestone units that contain reservoirs within the Lansing–Kansas City interval.

(1) [Heeb] — Subsurface elevation of the Heebner Shale, Upper Pennsylvanian

(2) [K-BP] — Thickness of the rock sequence from the base of the K-zone to the base of the Pennsylvanian rocks

(3) [I] — Thickness of the I-zone

(4) [Por H] — Thickness of porous carbonates in the H-zone

(5) [Por I] — Thickness of porous carbonates in the I-zone

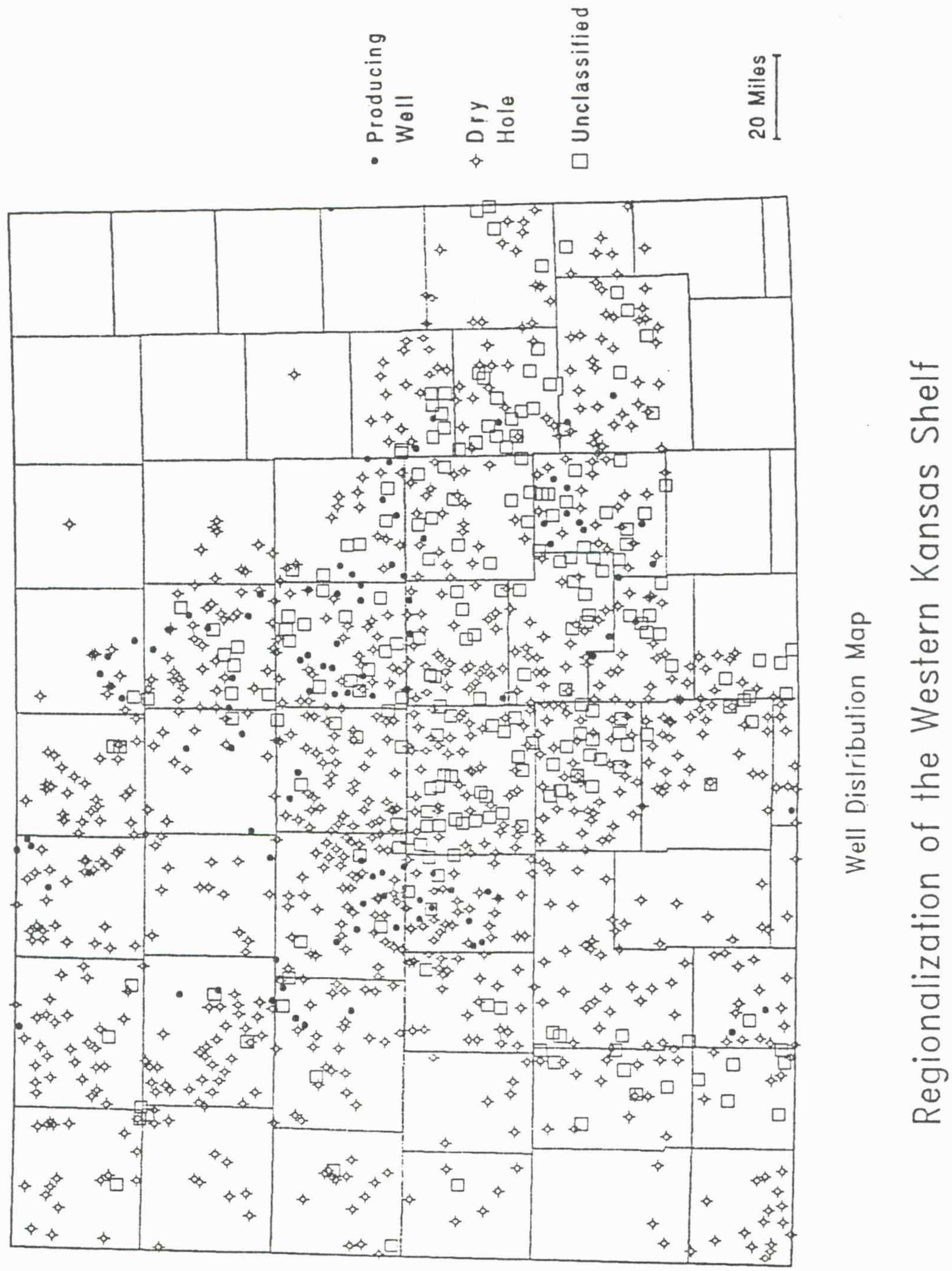

Figure 2. Locations of exploratory holes drilled in western Kansas used in this study. Symbols indicate productive status.

(6) [Ga J] — Maximum gamma-radiation response in marine shale of the J-zone.

Kriging Bayesian Probabilities

The experimental values $p[i \mid f(r_j)]$, $i \in \{\text{oil}, \text{dry}\}, j \in \{1,\ldots,1245\}$ of the regionalized variable were calculated by eq. (12) as discriminant scores produced by the software package SPSS^{X}™. Because only two classes ($p[\text{oil} \mid f(r)] + p[\text{dry} \mid f(r)] = 1$) are used, it is sufficient to consider only one variable. We focus here on the variable $p[\text{oil} \mid f(r)]$.

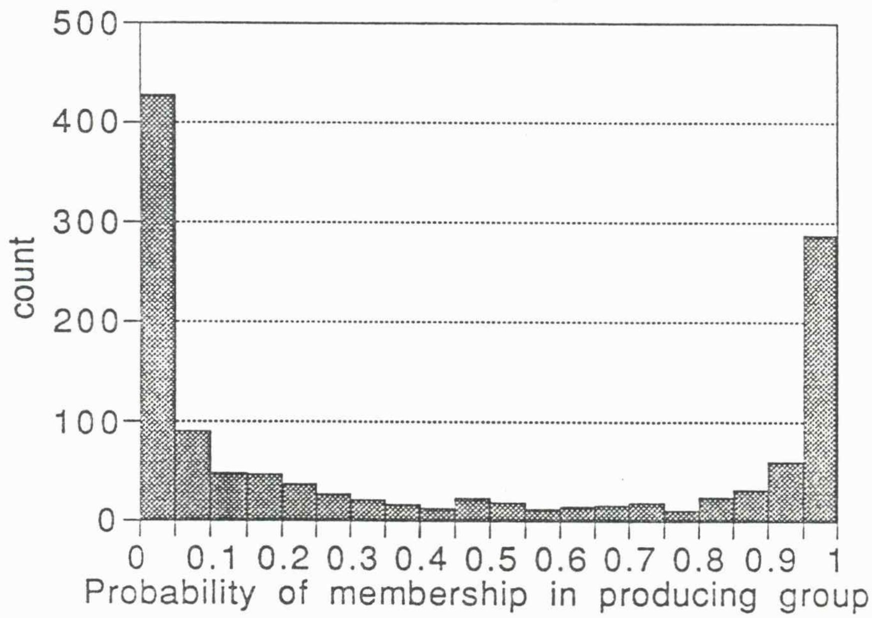

Figure 3. Frequency distribution of Bayesian probability $p[\text{oil} \mid f(r_i)]$, $i \in \{1,\ldots,1245\}$ of membership in the producing class.

Figure 3 shows the frequency distribution of $p[\text{oil} \mid f]$. The U-shape of the distribution function is caused by Bayes' transformation. Because in general sedimentary rocks in western Kansas strike northwest–southeast, the spatial correlation structure of the variable was expected to be anisotropic. The form of an experimental semivariogram in the northeast–southwest direction (Fig. 4) confirms this assumption. A northwest–southeast semivariogram is interpreted as reflecting the spatial correlation of the stochastic component and was used for the determination of model parameters.

Universal kriging with a 1st-order polynomial drift was used to produce the map shown in Figure 5. Negative estimates, as well as estimates greater than 1.0, result from the U-shaped frequency distribution of probability values (Fig. 3). The corresponding kriging variance is shown as a map in Figure 6.

Because of the reciprocal relationship $p[\text{dry} \mid f(r)] = 1 - p[\text{oil} \mid f(r)]$, the regionalization can be determined directly from the grid of the map in Figure 5. Regions with $p[\text{oil} \mid f(r)] > 0.5$ are assigned to class B_{oil}, or if $p[\text{oil} \mid f(r)] < 0.5$, they are assigned to class B_{dry}.

Kriging Mahalanobis' Distances

The decision procedure based on kriged Bayesian probabilities does not consider the uncertainty of interpolation due to the spatial distribution of observations. As an alternative to this procedure, we suggest that Mahalanobis' distances (eq. 11) be interpolated by treating the random vector of distances as a K-dimensional regionalized variable:

$$\Delta(r) = \begin{pmatrix} d_1^2(r) \\ d_2^2(r) \\ \vdots \\ d_K^2(r) \end{pmatrix}, \quad \forall r \in R. \tag{23}$$

Realizations of this variable are given by transformation of the measurement data at the well locations

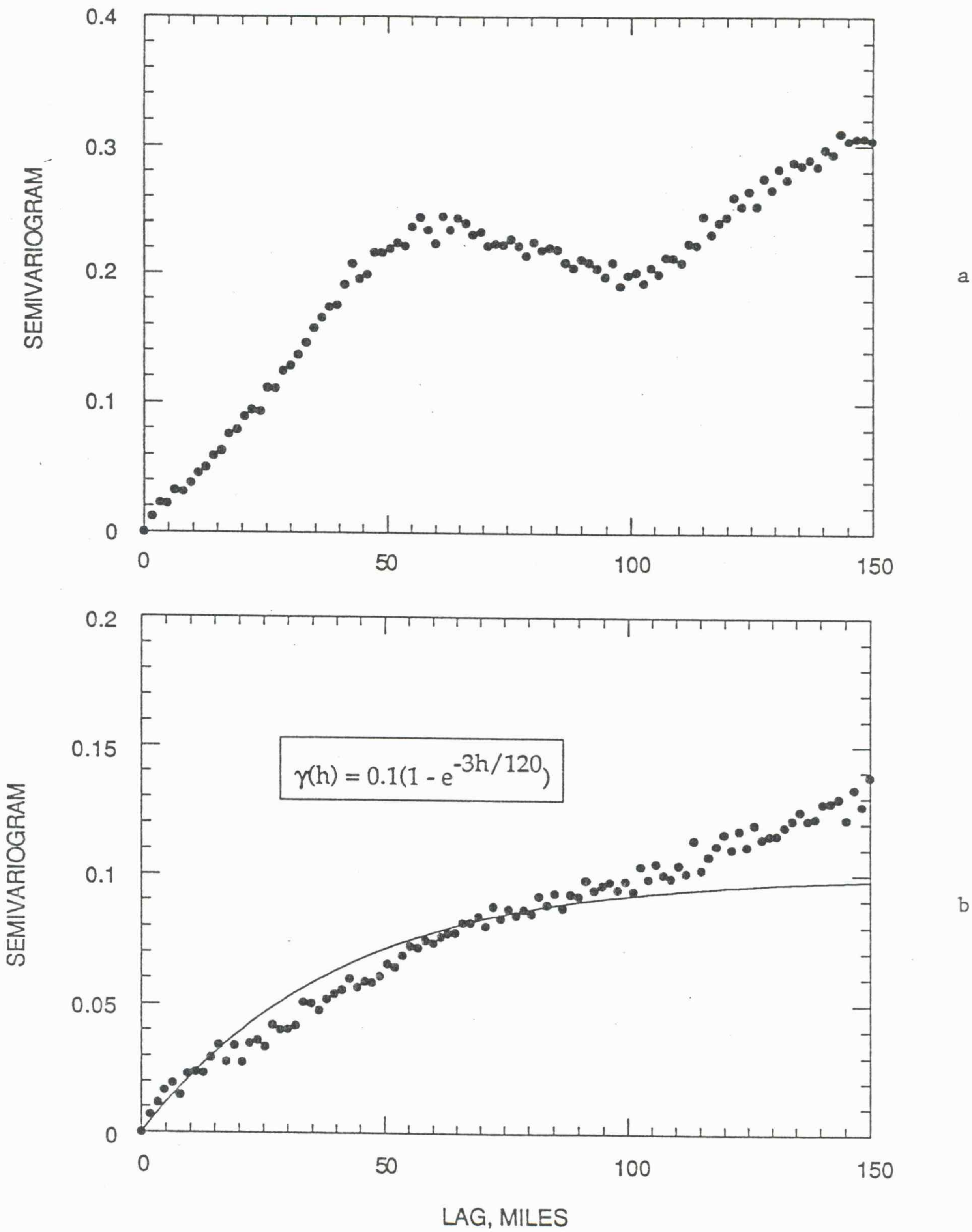

Figure 4. Semivariogram of $p[\text{oil} \mid f(r)]$: (a) experimental semivariogram in the northeast–southwest direction; (b) experimental semivariogram in the northwest–southeast direction. Line indicates exponential model with the given parameters.

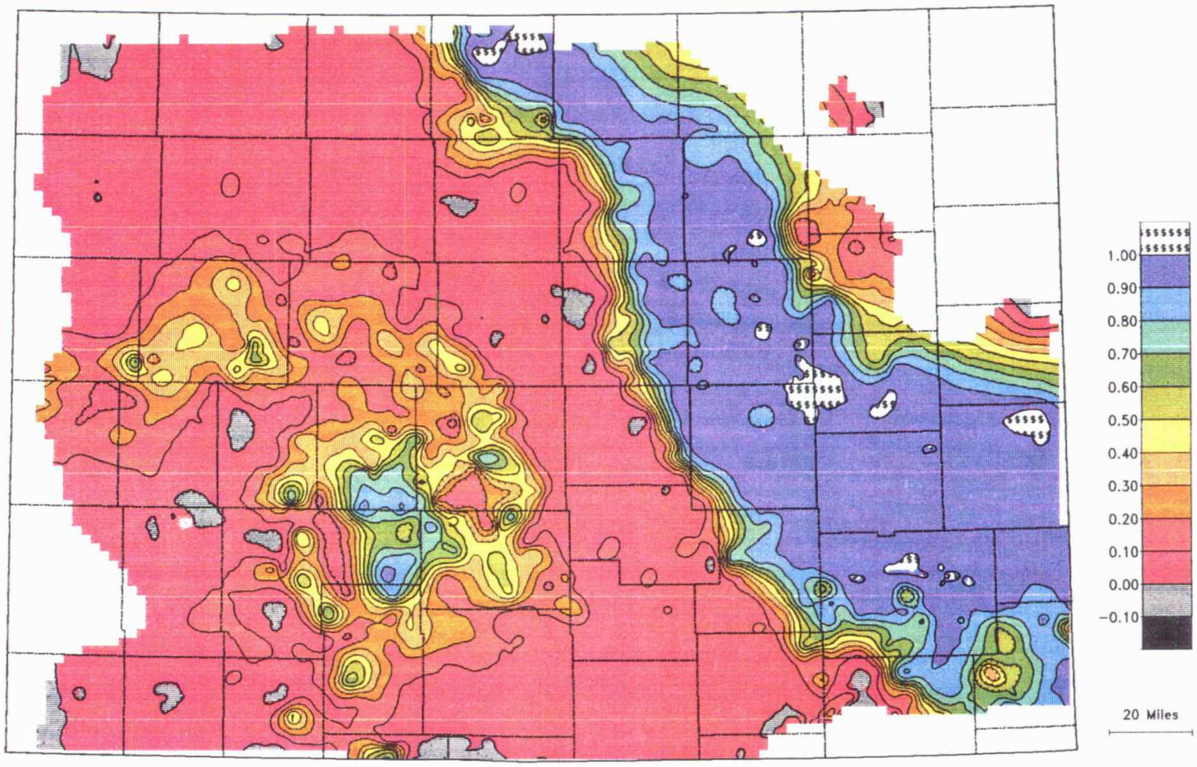

Figure 5. Map of Bayesian probability of membership in the producing class estimated using universal kriging of the variable $p[\text{oil} \mid f(r)]$.

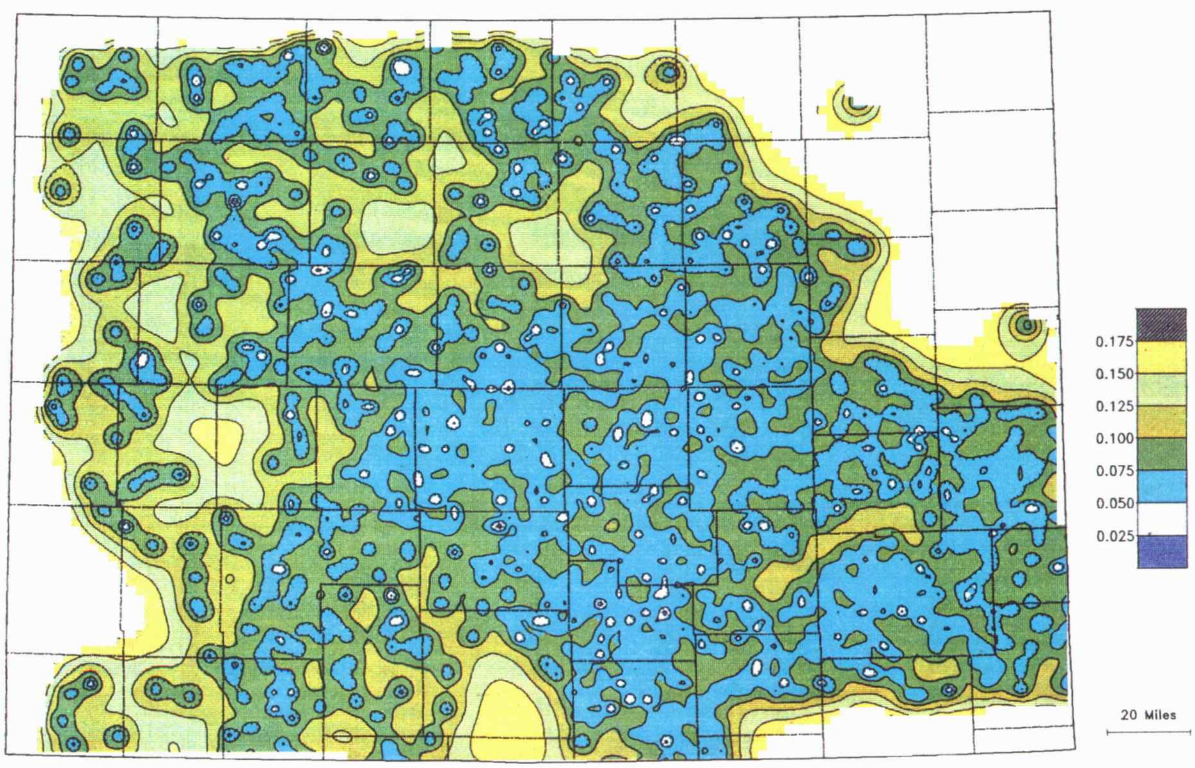

Figure 6. Variance of kriged estimates of Bayesian probability of membership in the producing class $p[\text{oil} \mid f(r)]$.

$$d_i^2\left(b(r_j)\right) = \left(x^G\left(b(r_j)\right) - m_i^*\right)' S_i^{-1}\left(x^G\left(b(r_j)\right) - m_i^*\right), \forall i \in I, \quad j \in \{1,\ldots,N\}, \tag{24}$$

The distribution of Mahalanobis' distances $d_i^2\left(b(r_j)\right)$ is nonsymmetrical because, if X has a normal distribution, the distances follow a χ^2-distribution with n degrees of freedom, where n is the number of variables. The frequency distributions of $d_{\text{oil}}^2(r)$ and $d_{\text{dry}}^2(r)$ for the western Kansas data are shown by Figure 7. These distances are not directly comparable because of differences between the estimated covariance matrices. For this reason the distances in discriminant analysis usually are corrected by the logarithm of the determinant of the covariance matrices (eqs. 10 and 12) as shown by Tatsuoka (1971).

To achieve unbiased kriging and a classification rule that incorporates the kriging variances, the distances must be standardized to Gaussian form by a normalization transformation Φ. Several equivalent methods are recommended for this purpose by Journel and Huijbregts (1978), Journel (1986), Hohn (1988), and Suro-Pérez and Journel (1991). The normally transformed variables $\hat{d}_i^2(r) = \Phi\left(d_i^2(r)\right)$ are assumed to be a regionalized variable whose drift can be modeled locally by a polynomial (eq. 15) and whose stochastic component can be described by a semivariogram (eq. 16). The experimental distances given by eq. (24) can be transformed by

$$\hat{d}_i^2\left(b(r_j)\right) = \Phi\left(d_i^2\left(b(r_j)\right)\right), \tag{25}$$

The results are transformed distance vectors

$$\hat{\Delta}\left(b(r_j)\right) = \begin{pmatrix} \hat{d}_1^2\left(b(r_j)\right) \\ \hat{d}_2^2\left(b(r_j)\right) \\ \vdots \\ \hat{d}_K^2\left(b(r_j)\right) \end{pmatrix}, \quad j \in \{1,\ldots,N\}. \tag{26}$$

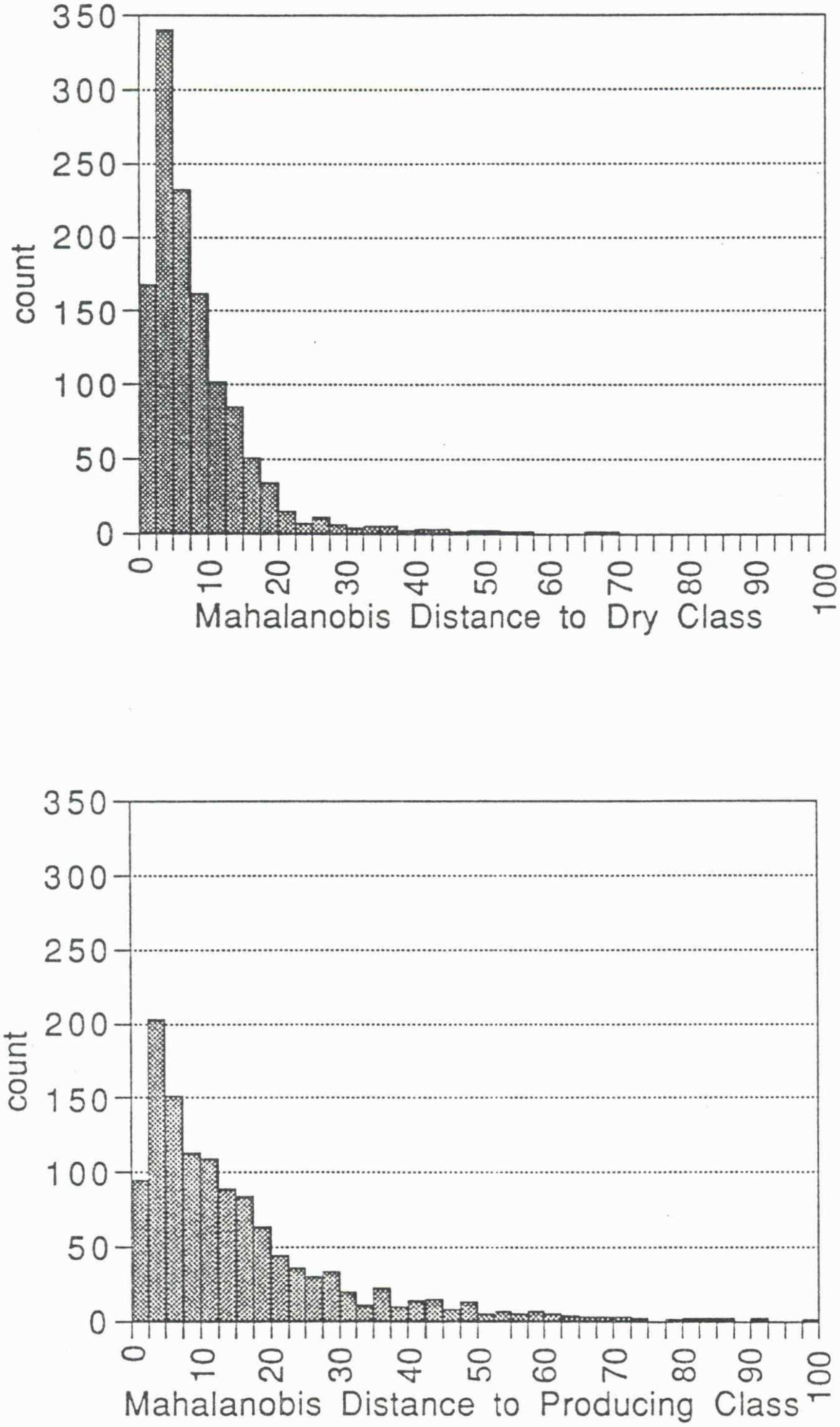

Figure 7. Frequency distributions of (a) Mahalanobis' distances $d^{2*}_{dry}(r)$ to the centroid of the dry class; and (b) Mahalanobis' distances $d^{2*}_{oil}(r)$ to the centroid of the producing class.

For a point $r_e \in R$ where no well has been drilled and therefore no well log measurements are available, a linear combination of the transformed distance values in the neighborhood J

$$\hat{\Delta}^*(r_e) = \begin{pmatrix} \hat{d}_1^{2^*}(r_e) \\ \hat{d}_2^{2^*}(r_e) \\ \vdots \\ \hat{d}_K^{2^*}(r_e) \end{pmatrix} \tag{27}$$

can be determined by kriging (eq. 17).

The location of a hypothetical well or drilling prospect at a point r_e where no logs are available, but where estimates $\hat{d}^{2^*}(r_e)$ have been interpolated, will be allocated to class Z', following the principle of minimal distance, $\min_{i \in I} \hat{d}_i^{2^*}(r_e)$, expressed by Tatsuoka (1971). By considering the kriging variance, the probability of correct classification can be calculated.

It is necessary to assume that interpolation errors are normally distributed. For a two-class situation, there are two estimated distances, so there are two normal error distributions with parameters $\mu_i = \hat{d}_i^{2^*}(r_e)$, $\sigma_i^2 = \sigma_i^2(r_e)$ and $\mu_j = \hat{d}_j^{2^*}(r_e)$, $\sigma_j^2 = \sigma_j^2(r_e)$. A point

$$t_{ij} = \hat{d}_i^{2^*}(r_e) + \bar{d}_{ij} \tag{28}$$

can be determined such that

$$\bar{d}_{ij} = \left[\hat{d}_j^{2^*}(r_e) - \hat{d}_i^{2^*}(r_e)\right] \sigma_i(r_e) / \left[\sigma_i(r_e) + \sigma_j(r_e)\right]. \tag{29}$$

The probabilities

$$p\left[\hat{d}_i^{2^*}(r_e) < t_{ij}\right] = \left(1 / \sigma_i(r_e)\sqrt{2\pi}\right) \int_{-\infty}^{t_{ij}} \exp\left\{-\tfrac{1}{2}\left(\frac{z - \hat{d}_i^{2^*}(r_e)}{\sigma_i(r_e)}\right)^2\right\} dz \tag{30}$$

and

$$p\left[\hat{d}_j^{2^*}(r_e) > t_{ij}\right] = \left(1/\sigma_j(r_e)\sqrt{2\pi}\right)\int_{t_{ij}}^{+\infty} \exp\left\{-\tfrac{1}{2}\left(\frac{z-\hat{d}_j^{2^*}(r_e)}{\sigma_j(r_e)}\right)^2\right\} dz \tag{31}$$

are equal,

$$p\left[\hat{d}_i^{2^*}(r_e) < t_{ij}\right] = p\left[\hat{d}_j^{2^*}(r_e) > t_{ij}\right]. \tag{32}$$

Obviously it holds that if $\hat{d}_i^{2^*}(r_e) < \hat{d}_j^{2^*}(r_e)$,

$$p\left[\hat{d}_i^{2^*}(r_e) < t_{ij}\right] > p\left[\hat{d}_j^{2^*}(r_e) < t_{ij}\right] \tag{33}$$

and if $\hat{d}_i^{2^*}(r_e) = \hat{d}_j^{2^*}(r_e)$,

$$p\left[\hat{d}_i^{2^*}(r_e) < t_{ij}\right] = p\left[\hat{d}_j^{2^*}(r_e) < t_{ij}\right] = 0.5. \tag{34}$$

Therefore, the probability of the correct allocation of a geological body $b(r_e)$ to class B_i and the rejection of membership in class B_j under the condition that the estimates show the relation $\hat{d}_i^{2^*}(r_e) < \hat{d}_j^{2^*}(r_e)$ and kriging variances exist can be expressed by $p\left[\hat{d}_i^{2^*}(r_e) < t_{ij}\right]$.

The probability given by eq. (30) can be calculated from the equation for the standardized normal distribution

$$p\left[\hat{d}_i^{2^*}(r_e) < t_{ij}\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d'_{ij}} \exp\left(-\tfrac{1}{2}Z^2\right) dz \tag{35}$$

with

$$d'_{ij} = \bar{d}_{ij} / \sigma_i(r_e) = \left[\hat{d}_j^{2*}(r_e) - \hat{d}_i^{2*}(r_e)\right] / \left[\sigma_i(r_e) + \sigma_j(r_e)\right]. \tag{36}$$

In the more general case that more than two classes exist, and $\hat{d}_i^{2*}(r_e) = \min_j \hat{d}_j^{2*}(r_e)$, the probability that an allocation of $b(r_e)$ to class B_i is correct can be expressed by the following equation:

$$p\left[i \mid \hat{\Delta}^*(r_e)\right] = \prod_{j \in I \setminus i} p\left[\hat{d}_i^{2^*}(r_e) < t_{ij}\right], \forall i \in I. \tag{37}$$

Equations (35) or (37) may replace the Bayesian expression commonly used in discriminant analysis. The allocation to one of the classes of partition Z is based on the principle of maximum probability.

$$b(r_e) \in B_i \Leftrightarrow p\left[i \mid \hat{\Delta}^*(r_e)\right] = \max_j p\left[j \mid \hat{\Delta}^*(r_e)\right]. \tag{38}$$

For the western Kansas data set, the Mahalanobis' distances $d_{\text{dry}}^2(r)$ and $d_{\text{oil}}^2(r)$ between each drill hole and the multivariate centroids of class B_{dry} and B_{oil} were calculated by eq. (35), using the mean vectors and covariance matrices given in Table 1. Figure 7 shows the frequency distribution of the unstandardized distances. The standardization method described by Journel and Huijbregts (1978) was used to transform this distribution into the frequency distribution of standardized distances shown in Figure 8. The two standardized distances are treated as composite variables that characterize each well location. Experimental semivariograms were estimated for the distance to class B_{dry} and the distance to class B_{oil} and are shown in Figure 9. The pronounced drift evident in the northeast–southwest semivariogram for the distance to class B_{oil} is a response to the Central Kansas Uplift, a large anticlinal feature that strikes north–northwest across the eastern part of the area. The Central Kansas Uplift was active throughout most of Paleozoic time and influenced sedimentary deposition; over its crest, most rock units are thin, relatively clean, and structurally high. The Central Kansas Uplift is also

Table 1. Experimental mean vectors and covariance matrices for six geological variables calculated for oil producing and dry wells in western Kansas.

Means

Variables	B_{dry}	B_{oil}	Dimension
Heeb	-1.16862E+03	-1.12877E+03	[feet]
K-BP	3.94781E+02	1.49952E+02	[feet]
I	2.09211E+01	1.74628E+01	[feet]
PorH	2.55348E+00	2.64386E+00	[feet]
PorI	2.73797E+00	1.60563E+00	[feet]
GaJ	1.74182E+02	1.69590E+02	[API]

Covariance Matrix S_{dry}

Heeb	K-BP	I	PorH	PorI	GaJ
1.47709E+05	5.02166E+04	-4.05434E+02	-2.15052E+02	-2.00492E+02	-6.36431E+02
5.02166E+04	3.92457E+04	1.53115E+03	6.98658E+01	2.00834E+02	-9.30886E+01
-4.05434E+02	1.53115E+03	2.00110E+02	1.14922E+01	3.63233E+01	1.20652E+02
-2.15052E+02	6.98658E+01	1.14922E+01	1.13479E+01	3.84000E+00	2.71851E+00
-6.36431E+02	-9.30886E+01	1.20652E+02	2.71851E+00	3.05859E+01	3.55428E+03

Covariance Matrix S_{oil}

Heeb	K-BP	I	PorH	PorI	GaJ
3.10636E+04	5.88609E+03	3.14115E+02	-4.35405E+01	3.46002E+00	1.96993E+03
5.88609E+03	1.08433E+04	7.08805E+02	8.04505E+01	1.73055E+01	3.01492E+03
3.14115E+02	7.08805E+02	9.80999E+01	7.64700E+00	6.83206E+00	1.78872E+02
-4.35405E+01	8.04505E+01	7.64700E+00	1.06652E+01	9.19752E-01	2.20836E+00
3.46002E+00	1.73055E+01	6.83206E+00	9.19752E-01	6.23529E+00	1.57350E+01
1.96993E+03	3.01492E+03	1.78872E+02	2.20836E+00	1.57350E+01	4.93130E+03

a preferred habitat for oil and gas, so its presence is strongly reflected in the geological characteristics of class B_{oil}, which contains most of the producing wells.

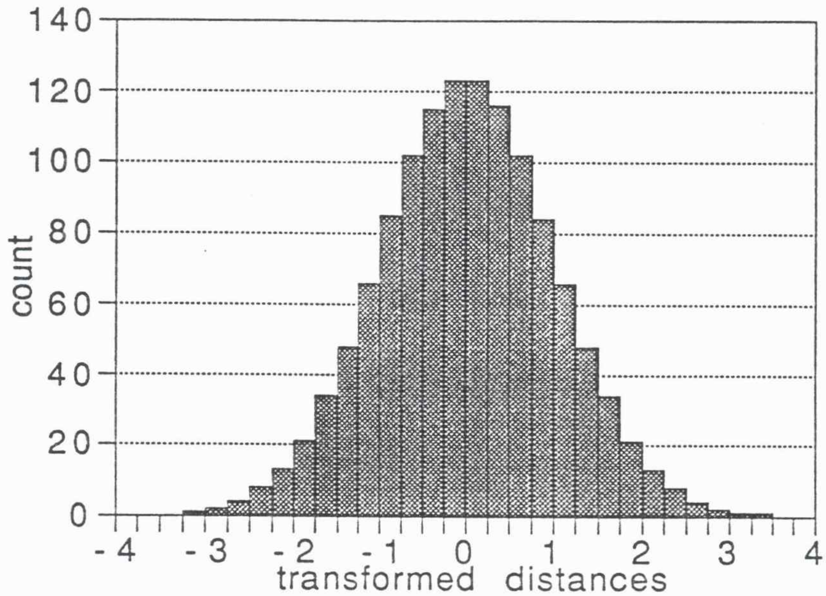

Figure 8. Frequency distribution of normalized Mahalanobis' distances $d_i^{2^*}(r), i \in \{\text{dry}, \text{oil}\}$.

Parallel to the Central Kansas Uplift, the distance measures are free of drift, as indicated by the semivariograms. The experimental semivariograms in the northwest–southeast direction were modeled by nested exponential functions, which are shown on the plots. Because of drift in the variables $\hat{d}^2_{\text{dry}}(r)$ and $\hat{d}^2_{\text{oil}}(r)$, universal kriging was used for interpolation and the local drift was modeled by a first-order polynomial.

Figure 10 is a contour map of $\hat{d}^{2^*}_{\text{dry}}(r)$, the estimated value of the standardized Mahalanobis' distance to the centroid of the dry class. The map grid contains over 13,000 grid nodes, each estimated by universal kriging of the values of $\hat{d}^2_{\text{dry}}(b(r))$ that have been calculated for each well location. Figure 11 is an equivalent contour map of $\hat{d}^{2*}_{\text{oil}}(r)$, the estimated Mahalanobis' distance to the centroid of the producing class.

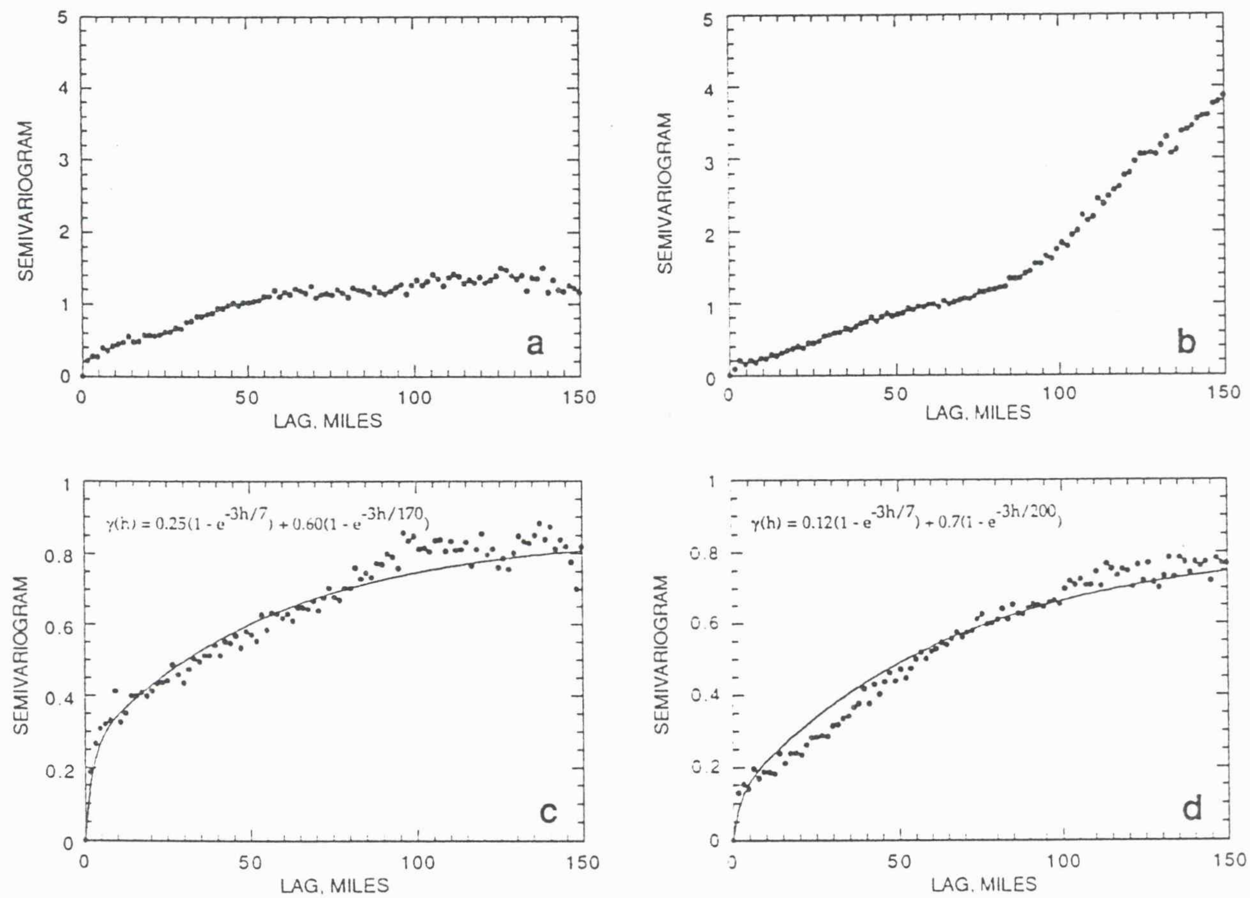

Figure 9. Variograms of Mahalanobis' distances. Both distance variables have been normalized ($\bar{x} = 0.00$, $s = 1.0$). (a) Distance to B_{dry}, measured in NE–SW direction. (b) Distance to B_{oil}, measured in NE–SW direction. (c) Distance to B_{dry}, measured in NW–SE direction. (d) Distance to B_{oil}, measured in NW–SE direction. Lines in (c) and (d) are nested exponential models whose coefficients are given on the variograms.

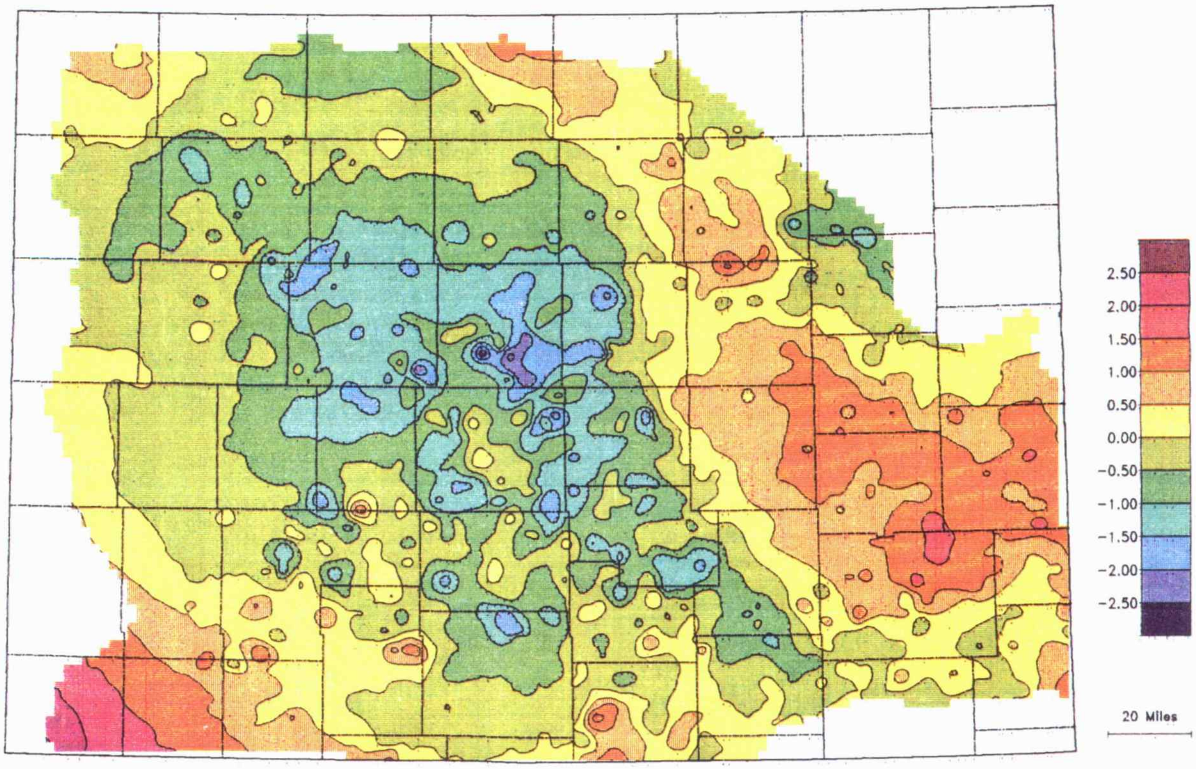

Figure 10. Contour map of Mahalanobis' distances $\hat{d}^{2*}_{\mathrm{dry}}(r)$ to B_{dry}, estimated using universal kriging. Distance values have been normalized.

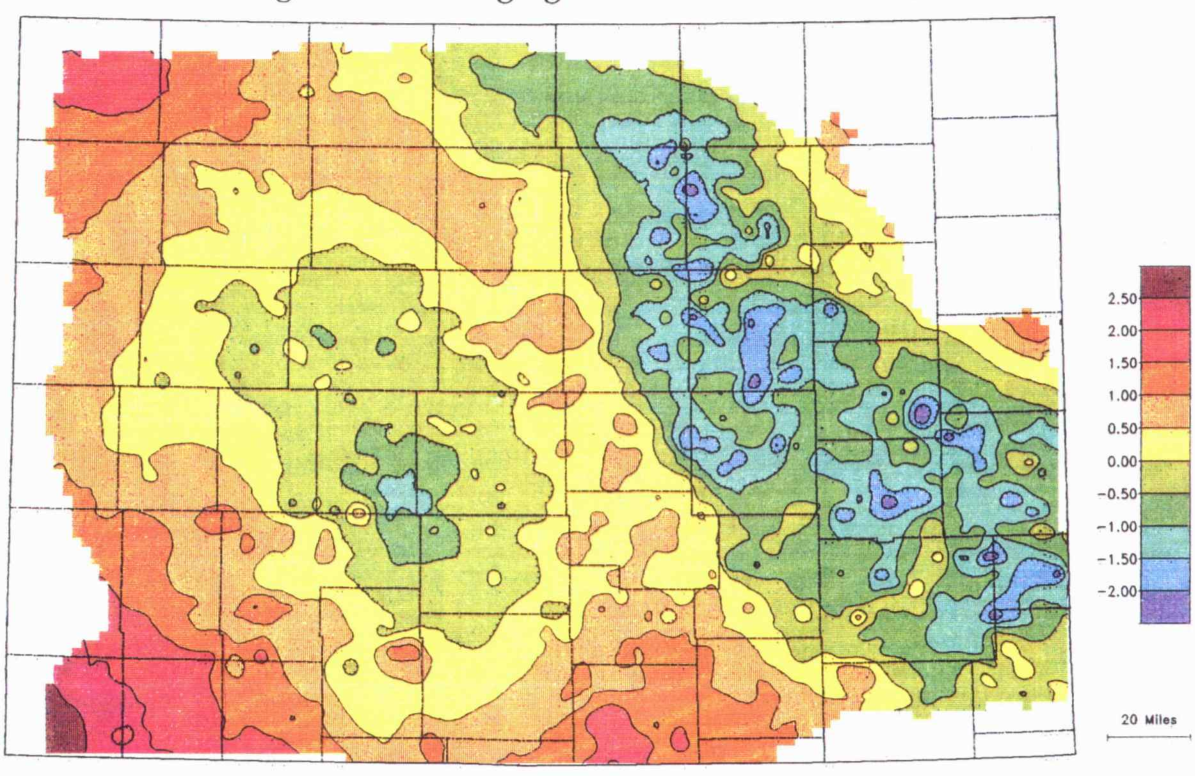

Figure 11. Contour map of Mahalanobis' distances $\hat{d}^{2*}_{\mathrm{oil}}(r)$ to B_{oil}, estimated using universal kriging. Distance values have been normalized.

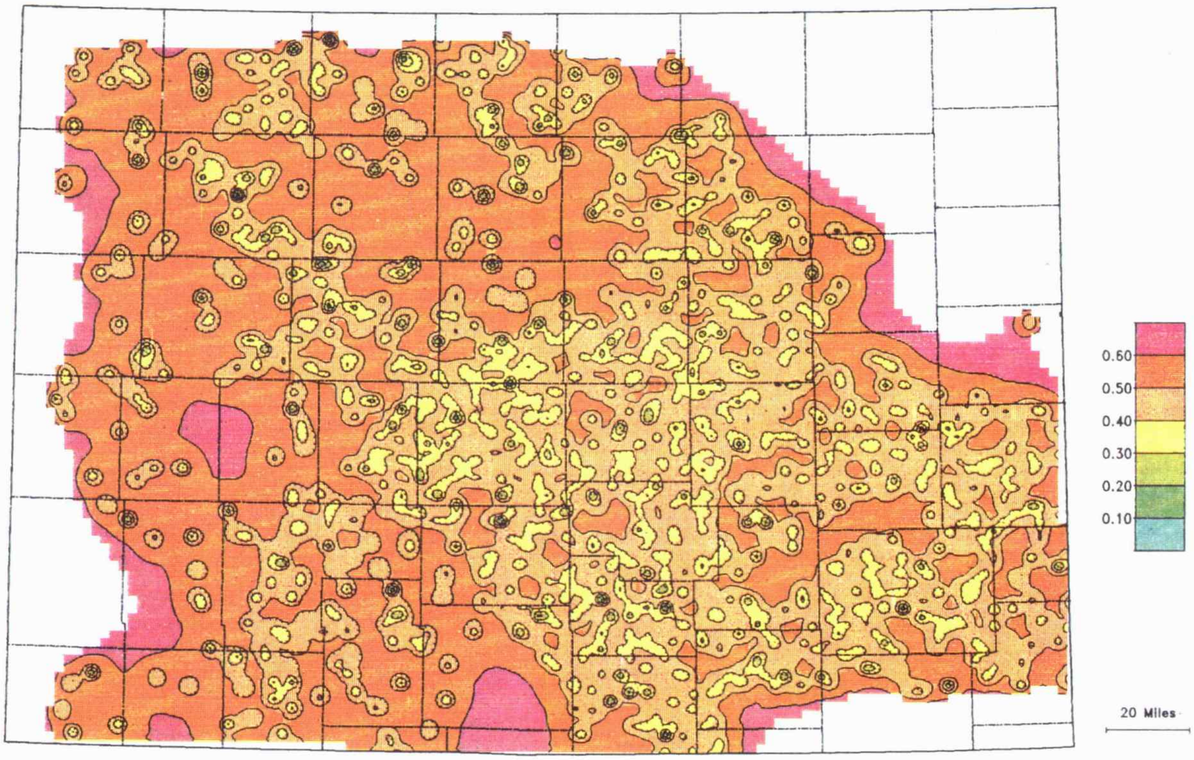

Figure 12. Kriging variance for kriged estimates of Mahalanobis' distance $\hat{d}^{2^*}_{\text{dry}}(r)$ to B_{dry}.

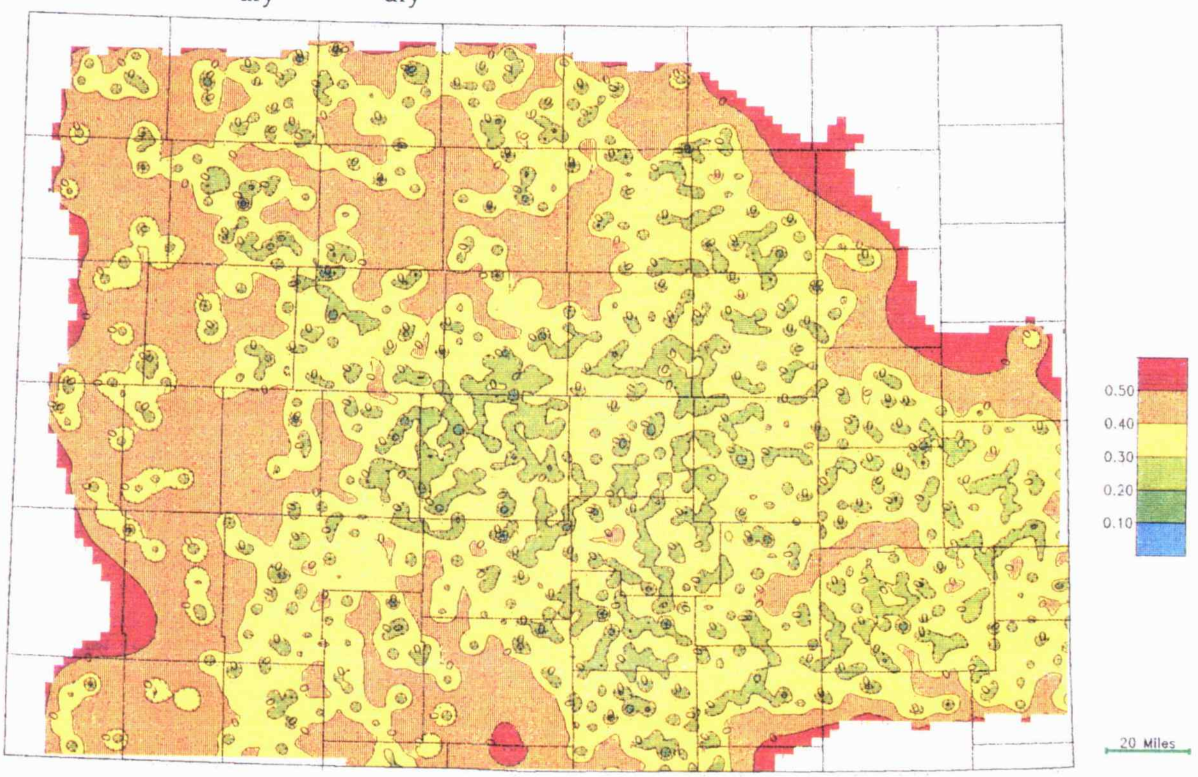

Figure 13. Kriging variance for kriged estimates of Mahalanobis' distance $\hat{d}^{2^*}_{\text{oil}}(r)$ to B_{oil}.

At every node, the kriging estimation variance also has been calculated for both variables and is shown in the form of contour maps in Figures 12 and 13. Values from these four map grids can be used in eq. (35) to estimate the probabilities that each grid node location belongs either to class B_{dry} or to class B_{oil}. Since the probabilities are estimated at all grid nodes, the results also can be displayed as a contour map, as in Figure 14. This map expresses the probability that an exploratory hole will be classified as belonging to B_{oil}.

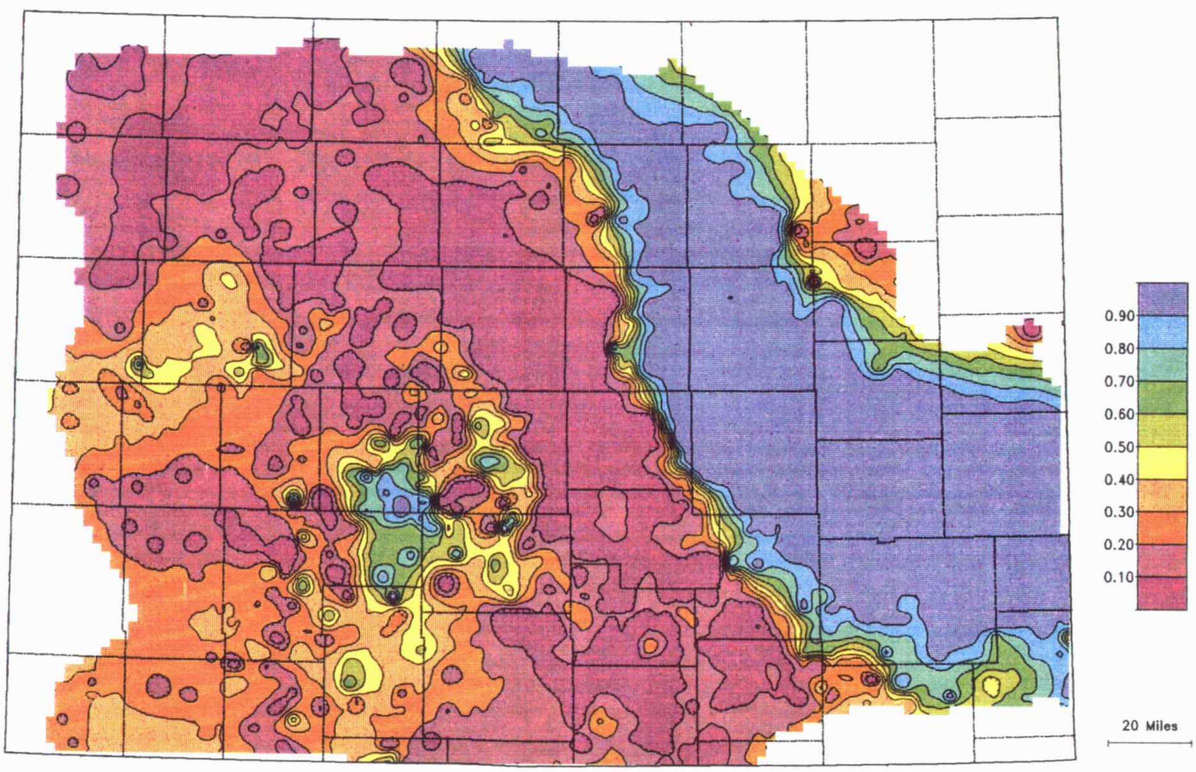

Figure 14. Probability that a well drilled in the study area will be classified as belonging to B_{oil}, based on Mahalanobis' distances and kriging variances.

Conclusions

Two different methods have been developed for estimating class membership of elementary geological bodies at unsampled locations. The advantage of the second method—the kriging of normalized Mahalanobis' distances and calculating the probabilities that the distances are minimal—is that the uncertainty of the interpolation is taken directly into account.

Comparison of results from the different methods demonstrates that both lead to the same general results. In the two-class problem, in the decision range close to 0.5 probability, the results are directly comparable. Differences are noticeable, however, in low and high probability ranges. These differences are caused by the different standardization methods used. Normalizing Mahalanobis' distances results in sensitive probability estimations at the upper and lower boundary of the probability range. In particular, the nonsymmetric shape of the frequency distributions close to these boundaries may result in negative kriging estimates of probabilities, or in estimates of probabilities greater than 1.0.

In addition, probability kriging does not consider the uncertainty in the estimation caused by the distribution of values in the neighborhood of the location to be predicted. Another advantage of normalized distance kriging is that the computational procedure is simpler than that of probability kriging. For these reasons, for multivariate decision-making problems in geology based on spatial interpolation, **normalized distance kriging** is recommended.

Acknowledgments

The authors thank the Kansas Geological Survey and the German Science Foundation for support of this research. Dr. Lynn Watney of the Kansas Geological Survey must be thanked for providing the data, and

Survey staff members J.A. DeGraffenreid, R.J. Sampson, and J.C. Wong provided technical assistance.

References

Fisher, R.A., 1936, The use of multiple measurements in taxonomic problems: *Annals of Eugenics*, no. 67, p. 179-188.

Harff, J., and Davis, J., 1990, Regionalization in geology by multivariate classification: *Mathematical Geology*, v. 22, p. 573-588.

Harff, J., Davis, J. C., Watney, L., Bohling, J., and Wong, J. C., 1989, Regionalization of western Kansas based on multivariate classification of stratigraphic data from oil wells: Open-File Rept. 89-21, Kansas Geol. Survey, Univ. Kansas, Lawrence, Kansas, 26 p., 9 figs., 26 maps.

Harff, J., Davis, J. C., and Olea, R.A., 1991, Quantitative assessment of mineral resources with an application to petroleum geology: *Journal of Nonrenewable Resources*, v. 1, no. 1 (in press).

Harff, J., Davis, J. C., Olea, R.A., and Watney, L., 1991, Three-dimensional regionalization and modeling for sedimentary basin analysis: Open-File Rept. 91-00, Kansas Geol. Survey, Univ. Kansas, Lawrence, Kansas (in prep.).

Haslett, J., 1989, Geostatistical neighborhoods and subset selection, in Armstrong, M. (Ed.), *Geostatistics, Proc. 3rd Geostat. Congress*, 5-9 Sept. 1988, Avignon, France: Kluwer Acad. Publ., Dordrecht, p. 569-578.

Hohn, M.E., 1988, *Geostatistics and Petroleum Geology*: Van Nostrand Reinhold, New York, 264 p.

Journel, A.G., 1986, Geostatistics: Models and tools for the earth sciences: *Mathematical Geology*, v. 18, p. 119-140.

Journel, A.G. and Huijbregts, C.J., 1978, *Mining Geostatistics*: Academic Press, London, 600 p.

Myers, D.E., 1982, Matrix formulation of co-kriging: *Mathematical Geology*, v. 14, p. 249-257.

Suro-Pérez, V., and Journel, A.G., 1991, Indicator principal component kriging: *Mathematical Geology*, v. 23, no. 5, p. 759-788.

Tatsuoka, M.M., 1971, *Multivariate Analysis: Techniques for Educational and Psychological Research*: John Wiley & Sons, New York, 310 p.

Watney, W.L., 1984, Recognition of favorable reservoir trends in Upper Pennsylvanian cyclic carbonates in western Kansas; *in* Hyne, N.J., ed., *Limestones of the Mid-continent*: Tulsa Geological Society, Special Publication No. 2, p. 201-246.

Watney, W.L., Harff, J., Davis, J.C., Bohling, G. and Wong, J.C., n.d., Controls on petroleum accumulation in Upper Pennsylvanian cyclic shelf carbonates in Western Kansas, USA, interpreted by space modeling and multivariate techniques; *in* Harff, J. and Merriam, D., eds., *Computerized Basin Analysis*: Plenum Press, New York (in press).