Quantitative Assessment of Mineral Resources with an Application to Petroleum Geology

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

[1]Central Institute for Physics of the Earth, Telegrafenberg, O-1561 Potsdam, Germany, and [2]Kansas Geological Survey, 1930 Constant Ave., Lawrence, Kansas 66047, USA

The probability of occurrence of natural resources, such as petroleum deposits, can be assessed by a combination of multivariate statistical and geostatistical techniques. The area of study is partitioned into regions that are as homogeneous as possible internally while simultaneously as distinct as possible. Fisher's discriminant criterion is used to select geological variables that best distinguish productive from nonproductive localities, based on a sample of previously drilled exploratory wells. On the basis of these geological variables, each wildcat well is assigned to the production class (dry or producer in the two-class case) for which the Mahalanobis' distance from the observation to the class centroid is a minimum. Universal kriging is used to interpolate values of the Mahalanobis' distances to all locations not yet drilled. The probability that an undrilled locality belongs to the productive class can be found, using the kriging estimation variances to assess the probability of misclassification. Finally, Bayes' relationship can be used to determine the probability that an undrilled location will be a discovery, regardless of the production class in which it is placed. The method is illustrated with a study of oil prospects in the Lansing/Kansas City interval of western Kansas, using geological variables derived from well logs.

Key words:
Resource assessment
Classification
Regionalization
Risk appraisal
Hydrocarbons

Introduction

Predicting the occurrence of mineral or energy resources requires knowledge about the geology of the region and about the geological processes responsible for the formation and accumulation of these resources. Geological processes are not directly observable and, because of their extreme time scale, simulating the formation of resources would require an analogue of complicated physical and chemical conditions over millions of years. Fortunately, geological features that result from these processes are observable and measurable at different spatial scales, although in most investigations these features can be observed only at isolated points.

Specific modeling procedures have been developed to help overcome the inadequacies of our geological knowledge; increasingly, these models employ mathematical methods implemented on computers. As a consequence, the amount of data and information derived by modeling has increased and methods for integrating geological data have been developed. The models permit us to interpret complex sets of geological information to assess the "favorability" of a region for resources. Favorability may be expressed qualitatively by maps that show areas where resources may be found, or quantitatively by maps that show the probability of resource occurrence.

Agterberg (1989) developed a model for calculating the regional potential of an area, based on the integration of geological maps. However, his methodology does not consider spatial interpolation, which is especially important in petroleum geology, where most information comes from petrophysical logs for widely spaced exploratory holes.

Data from well logs are integrated to produce regional maps indicating the probability of occurrence of oil or gas. The method is based on a procedure (Harff and Davis, 1990) that combines the theory of classification of geological objects (Rodionov, 1981) with that of regionalized variables (Matheron, 1970; Journel and Huijbregts, 1978). The interpolation aspects of regionalization are emphasized. A case study of regional favorability for oil occurrence in the Pennsylvanian Lansing and Kansas City Groups in western Kansas illustrates the technique.

Targets and Predictors

Geographic locations in the area of investigation are symbolically indicated by R, denoting a set of longitude and latitude vectors $r \in R$. The production status X^E of wells must be related to geological variables X^G measured on the rock sequences $b_i \in B$, $i \in \{1, \ldots, N\}$, penetrated by wells at locations r_i. This relationship must then be inverted to predict the status of undrilled localities from inferred values of the geological properties. X^E is the "target" variable and X^G, which commonly is multidimensional, is the "predictor" variable. In oil exploration, X^E is derived from production records and may include initial production, cumulative production over a specified interval, or similar measures. X^G may be any geological, geophysical, or petrophysical property that can be measured at the individual boreholes. In other contexts, X^E may be ore grade, tonnage, water-flow rate, or other measures of productivity or worth.

Solving this problem presupposes knowledge of the relationship $X^E = f(X^G)$. Unfortunately, some of the relevant geological variables may not be measurable, nor is the relationship between X^G and X^E necessarily consistent from one geological region to another. For these reasons, the function f cannot be described in an explicit analytical form.

One way to overcome this difficulty is to explore the function f by statistical means, using a random sample $B' \subset B$ for which both geological variables and measures of productivity are available. Multivariate regression can be used appropriately for this purpose if both kinds of variables are measured quantitatively. The experimentally determined function f can be considered valid for the prediction of X^E at those locations where only geological variables are available.

If the correlation between X^G and X^E is weak, the regression between the two may be uninformative. In this circumstance, X^E may be ranked and the ranks used to subdivide the set B' into a partition Z^E consisting of classes or subsets:

$$Z^E = \{B_1^E, B_2^E, \ldots, B_K^E\}, \qquad B_i^E \subset B'. \tag{1}$$

In addition, a partition

$$Z^G = \{B_1^G, B_2^G, \ldots, B_K^G\}, \qquad B_i^G \subset B', \tag{2}$$

which is described only by geological variables, must be determined so that each predictor class $B_i^G \subset B'$, $\forall i \in I$, $I = \{1, \ldots, K\}$, can be allocated to one target class $B_i^E \subset B'$. Geological predictor variables must be selected so that the size of $B_i^E \cap B_i^G$, $\forall i \in I$, is a maximum. The discriminant function criterion of Fisher (1936) can be used to select these n geological variables. Determining the function f is now simplified to investigating the relation $Z^E = f(Z^G)$. We may express f as the conditional probabilities that a well assigned to a predictor class B_i^G is also a member of a target class B_j^E.

In our application, geological properties are measurements derived from petrophysical logs run in the wells, and predictor classes are based on oil productivity of the

wells. The rank partition Z^E is given by values $x_j^E, j \in \{1, \ldots, k\}$, of one discrete variable X^E measured on wells $b_i \in B', i \in \{1, \ldots, N\}$, for which the logs are available:

$$\begin{pmatrix} x^E(b_1) \\ x^E(b_2) \\ \vdots \\ x^E(b_N) \end{pmatrix}. \tag{3}$$

These values may express, for example, K ranges of average daily oil production. Each range represents a class $B_i^E \subset Z^E$. In the simplest case ($K = 2$), X^E simply expresses the alternatives of a dry hole or a producing well.

The division Z^G given by equation 2 is determined for an n-dimensional geological variable representing information derived from well log records b_i, and ordered as a matrix:

$$\begin{pmatrix} x_1^G(b_1), & \ldots, & x_n^G(b_1) \\ x_1^G(b_2), & \ldots, & x_n^G(b_2) \\ & \vdots & \\ x_1^G(b_N), & \ldots, & x_n^G(b_N) \end{pmatrix}. \tag{4}$$

These data represent a mixture

$$P(X^G) = \sum_{i \in I} p[X^G \mid i] \tag{5}$$

of K n-dimensional probability distributions, $p[X^G \mid i]$.

Z^G and the evaluations of the probability distributions $p[X^G \mid i]$ can be determined from the matrix in equation 4, either by supervised classification using discriminant techniques (Rodionov, 1981) or by unsupervised classification analyzing the structure of the point cloud given by the row vectors of equation 4 in n-dimensional space. Unsupervised classification was used by Harff and Davis (1990) in an application of hierarchical cluster analysis.

The mean vectors m_i and the covariance matrices Σ_i of each class can be estimated by m_i^* and S_i, using data $x^G(b_j)$ from the supervised or unsupervised classification of rows in the matrix (eq. 4). Under the condition $m_i^* \neq m_j^*$, $\forall i,j \in I$, $i \neq j$ for each class $B_i^G \subset Z^G$, the following classification rule is assumed to be valid:

$$B_i^G = \left\{ b_j \in B' : d_i^2(b_j) = \min_k d_k^2\left(d_k^2(b_j)\right) \right\} \tag{6}$$

This classification rule uses Mahalanobis' distance

$$d_i^2(b_j) = (x^G(b_j) - m_i^*)' S_i^{-1} (x^G(b_j) - m_i^*). \tag{7}$$

If the $p[X^G \mid i]$ follow a normal distribution, the a posteriori probability given by Bayes' equation (Tatsuoka, 1971) can be used instead of equation 6 in the determination of Z^G. It is also possible to use Mahalanobis' distance based on discriminant scores instead of the original variables.

The relation f between Z^G and Z^E is an expression of the conditional probability that a well belonging to predictor class B_i^G is also a member of target class B_j^E. This probability can be estimated by the Bayesian relation

$$p[E_j \mid G_i] = \frac{M(B_i^G \cap B_j^E)}{M(B_i^G)}, \qquad \forall i,j \in I, \tag{8}$$

where $M(\cdot)$ is the size of the set.

Interpolation

Equation 6 is assumed to be valid for extrapolation, which means that unclassified wells $b \in B \backslash B'$ can be assigned via the predictor classes to their appropriate target classes. However, such extrapolation is of limited interest; we wish, rather, to assess locations where no wells have been drilled and, consequently, no well logs are available. For this purpose, extrapolation must also involve spatial interpolation. Myers (1982) has described a generalized cokriging method for interpolating an n-dimensional geological variable. For practical reasons, this method may prove difficult to apply in many circumstances (Harff and Davis, 1990). In addition, classification by a discriminant function (eq. 6) assumes the observations are independent. This assumption is not valid for estimates made by spatial interpolation.

As an alternative, we suggest that Mahalanobis' distances (eq. 7) be interpolated. The random vector of distances is regarded as a k-dimensional regionalized variable:

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

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

$$d_i^2(b(r_j)) = (x^G(b(r_j)) - m_i^*)' S_i^{-1} (x^G(b(r_j)) - m_i^*), \quad \forall i \in I, \quad j \in \{1, \ldots, N\}. \tag{9}$$

The distribution of Mahalanobis' distances $d_i^2(b(r_j))$ is nonsymmetrical because, if X^G has a normal distribution, the distances (eq. 9) follow a χ^2-distribution with n degrees of freedom. These distances are not directly comparable because of differences among the estimated co-

variance matrices. For this reason the distances in discriminant analysis usually are corrected by an additive term consisting of the logarithm of the determinant of the covariance matrices (Tatsuoka, 1971).

For our purposes (unbiased kriging and a classification rule based on kriging variances), the distances are standardized to a Gaussian distribution by a normalization transformation Φ. Equivalent methods are recommended by Journel and Huijbregts (1978), Journel (1986), and Hohn (1988). The normally transformed variables $\hat{d}_i^2(r) = \Phi(d_i^2(r))$ are assumed to be regionalized variables with a deterministic drift $m(r)$ and a stochastic component $Y(r)$:

$$\hat{d}_i^2(r) = m(r) + Y(r), \qquad \forall r \in R, \qquad \forall i \in I.$$

Because a condition of stationarity is not required, the drift can be modeled locally by

$$\mathrm{E}[\hat{d}_i^2(r)] = m(r) = \sum_{j \in J} a_j f_j(r). \tag{10}$$

The spatial structure of the stochastic component is described by a semivariogram

$$\gamma_i(h) = \tfrac{1}{2}\mathrm{E}[(Y(r+h) - Y(r))^2]. \tag{11}$$

The experimental distances given by equation 9 can be transformed by

$$\hat{d}_i^2(b(r_j)) = \Phi(d_i^2(b(r_j))). \tag{12}$$

The results are transformed distance vectors

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

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,

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

can be determined by

$$\hat{d}_i^{2*}(r_e) = \sum_{j \in J} \lambda_j \hat{d}_i^2(b(r_j)), \qquad i \in I. \tag{13}$$

The λ_j are determined by solving the universal kriging system of equations using an appropriate model fitted to the experimental variogram. Under these conditions, the estimates given by equation 13 are unbiased. The corresponding kriging variance is denoted by $\sigma_i(r)$, $i \in I$.

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 a class Z^G, following the principle of minimal distance, $\min_{i \in I} \hat{d}_i^{2*}(r_e)$, expressed in equation 6. By using the kriging variance, the probability that a distance $\hat{d}_i^{2*}(r_e)$ is a minimum can be calculated.

It is necessary to assume the interpolation error is normally distributed. In a two-class situation, there is a pair of estimated distances, so there are two normal error distributions with parameters $\mu_i = \hat{d}_i^{2*}(r_e)$, $\sigma_i = \sigma_i(r_e)$ and $\mu_j = \hat{d}_j^{2*}(r_e)$, $\sigma_j = \sigma_j(r_e)$. The probability that $\mu_i = \hat{d}_i^{2*}(r_e)$ is smaller than $\mu_j = \hat{d}_j^{2*}(r_e)$ is given by the integral of error distribution i from $-\infty$ to $\bar{d}_{ij}$, *where* $\bar{d}_{ij}$ is the crossing point between the two distributions. This probability can be calculated from the equation for the standardized normal distribution

$$p[\hat{d}_i^{2*}(r_e) < \hat{d}_j^{2*}(r_e)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\bar{d}_{ij}} \exp\left(-\frac{1}{2}z\right) dz, \tag{14}$$

with

$$\bar{d}_{ij} = [\hat{d}_j^{2*}(r_e) - \hat{d}_i^{2*}(r_e)]/[\sigma_i(r_e) + \sigma_j(r_e)]. \tag{15}$$

If there are more than two classes, the probability that a distance $\hat{d}_i^{2*}(r_e)$ is the minimum within the vector of distances can be expressed by

$$p\left[\hat{d}_i^{2*}(r_e) = \min_j \hat{d}_j^{2*}(r_e)\right] = \prod_{j \in I \setminus i} p[\hat{d}_i^{2*}(r_e) < \hat{d}_j^{2*}(r_e)], \qquad \forall i \in I. \tag{16}$$

Equations 14 and 16 replace the Bayesian expression commonly used in discriminant analysis. For a prospect at location r_e, where estimates $\hat{\Delta}^*(r_e)$ and the corresponding kriging variances are available, equation 17 denotes the probability that the well belongs to one of the predictor classes $Z^G = \{B_i^G\}$, $i \in I$:

$$p[i \mid \hat{\Delta}^*(r_e)] = p\left[\hat{d}_i^{2*}(r_e) = \min_j \hat{d}_j^{2*}(r_e)\right], \qquad \forall i \in I. \tag{17}$$

A more detailed discussion of interpolation and classification will follow in a future report.

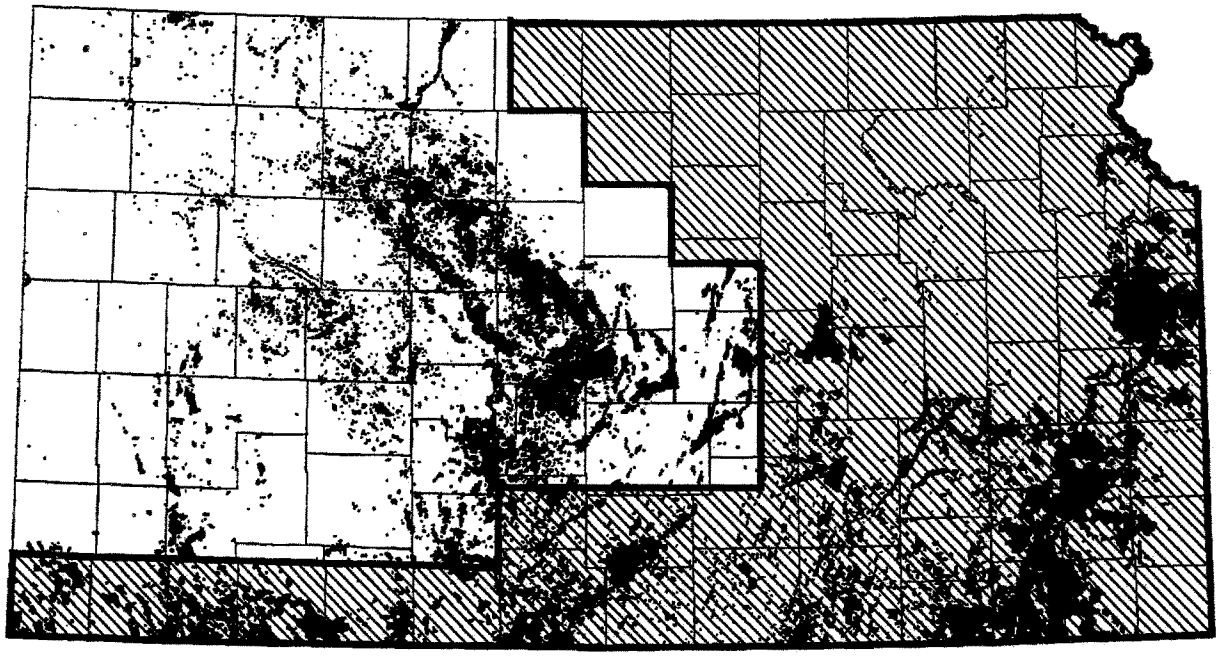


Figure 1. Oil fields in Kansas. Unruled area indicates the part of Kansas covered by this study.

Risk Assessment

For risk assessment in petroleum exploration, we seek a target class B_j^E that is characterized by the highest production rate. The corresponding predictor class B_i^G is given by

$$B_i^G \subset Z^G: p[E_j|G_i] = \max_k p[E_j|G_k].$$

For regional evaluation of exploration risk, the probability that a point should be assigned to the predictor class (given by eq. 17) must be multiplied by the conditional probability that a well is a producer, given that it is a member of the predictor class (eq. 8). That is,

$$p[G_i \wedge E_j|\hat{\Delta}^*(r_e)] = p[i|\hat{\Delta}^*(r_e)] \cdot p[E_j|G_i]. \quad (18)$$

The probability of a discovery at point r_e, if B_j^E is the target class, can be expressed as the equation

$$p[E_j(r_e)] = \sum_{i \in I} p[G_i \wedge E_j|\hat{\Delta}^*(r_e)]. \quad (19)$$

Exploration risk can be calculated as

$$p[\sim E_j(r_e)] = 1 - p[E_j(r_e)]. \quad (20)$$

Case Study

This example involves regionally classifying localities in the western Kansas shelf area and estimating the probabilities of discovering oil at these locations. The prospective interval in the study is the Lansing and Kansas City Groups of Pennsylvanian (Upper Carboniferous) age, consisting of interbedded marine limestones and shales deposited in a shallow epicontinental sea that covered the North American midcontinent in the late Paleozoic. 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.

Data were collected from 1,245 exploratory drill holes that penetrate the Lansing/Kansas City. Of these drill holes, 917 were dry, 99 produced oil in the Lansing/Kansas City interval, and 229 were unclassified, with unknown producing status. Well locations are posted on figure 2. Geological information, consisting of measure-

Table 1. Experimental mean vectors for six geological variables calculated for dry (class 1) and oil-producing (class 2) wells in western Kansas.

Variables	B_1^G	B_2^G	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
Porl	2.73797E+00	1.60563E+00	feet
GaJ	1.74182E+02	1.69590E+02	API

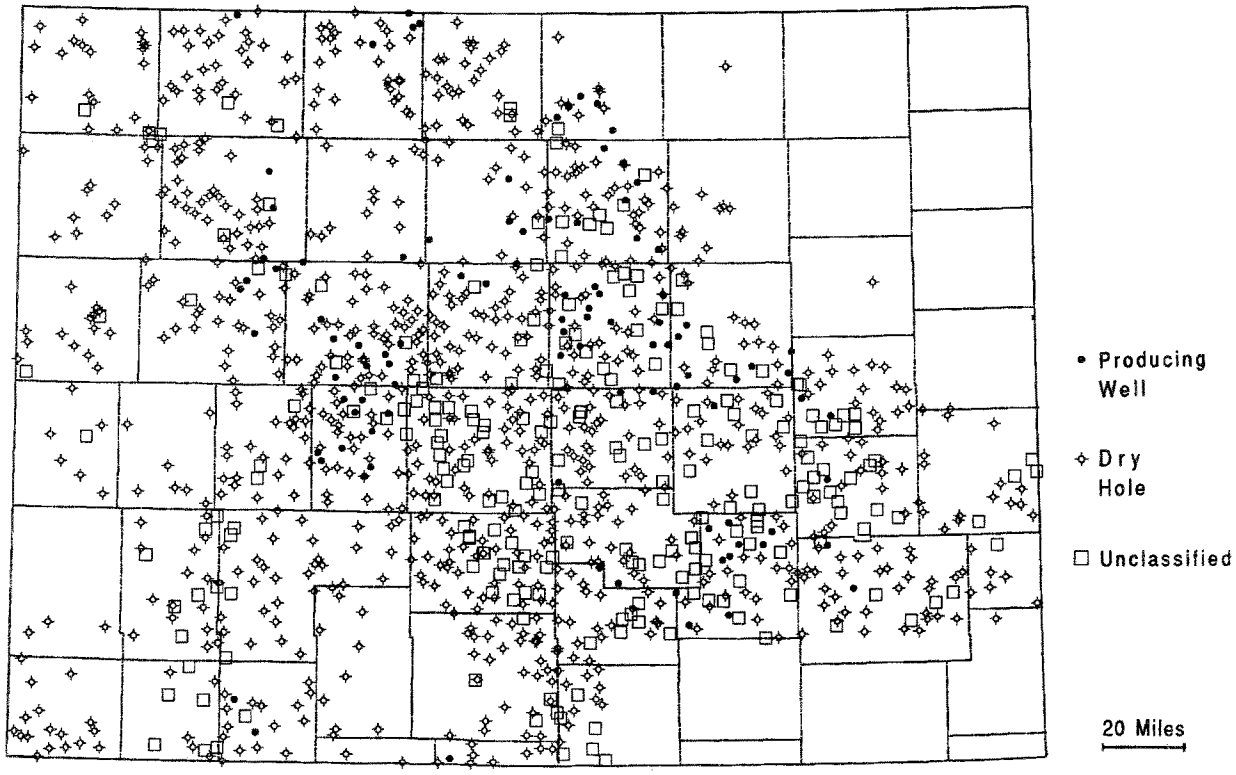


Figure 2. Locations of exploratory holes drilled in western Kansas used in this study. Symbols indicate productive status.

ments to the top of key formations, the thickness of selected stratigraphic intervals, and petrophysical properties averaged over selected intervals, was recorded on well logs from each hole. In this example, only two classes of productivity were used: dry ($x^E = 0$ bbl oil/day) or oil producing ($x^E > 0$ bbl oil/day). Z^E was determined as $\{B^E_{\text{oil}}, B^E_{\text{dry}}\}$, $b_i \in B^E_{\text{oil}}$, $i \in \{1, \ldots, 99\}$, $b_i \in B^E_{\text{dry}}$, $i \in \{1, \ldots, 917\}$.

Sixteen geological variables originally were recorded; preliminary studies indicated that some were either redundant or only weakly related to oil production. The discriminant function between the two groups was calculated, and Fisher's criterion was used to determine the effective discriminating variables. Six variables proved to have significant discriminating power (the terms "K-zone," "I-zone," etc., refer to specific limestone units within the Lansing/Kansas City interval that contain reservoirs):

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. PorH—thickness of porous carbonate in the H-zone
5. PorI—thickness of porous carbonate in the I-zone
6. GaJ—maximum gamma-radiation response in the marine shales of the J-zone.

A supervised class assignment of each drill hole to predictor classes $Z^G = \{B^G_1, B^G_2\}$, $b_i \in B^G_1$, $i \in \{1, \ldots, 628\}$, $b_i \in B^G_2$, $i \in \{1, \ldots, 388\}$, was determined based on its discriminant score. The vectors of means and covariance matrices of these variables for the two classes were calculated and are given in tables 1–3. Using the Bayesian relationship given in equation 8, the following conditional probabilities were estimated:

Table 2. Experimental covariance matrix S_1 for six geological variables calculated for dry wells in western Kansas.

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

Table 3. Experimental covariance matrix S_2 for six geological variables calculated for oil-producing wells in western Kansas.

Heeb	K-BP	I	PorH	Porl	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

$p[\text{oil}|1] = 0.35$, $p[\text{oil}|2] = 0.69$,
$p[\text{dry}|1] = 0.65$, $p[\text{dry}|2] = 0.31$.

The Mahalanobis' distances $d_1^2(r)$ and $d_2^2(r)$ between each drill hole and the multivariate centroids of class 1 and class 2 were calculated by equation 9, using the mean vectors and covariance matrices of tables 1–3. A standardization method described by Journel and Huijbregts (1978) was used. The two standardized distances are treated as composite variables that characterize each well location. Experimental semivariograms were estimated for the distance to class 1 and the distance to class 2 and are shown in figure 3. The pronounced drift evident in the northeast-southwest semivariogram for the distance to class 2 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 the crest of the Uplift, most rock units are thin, relatively clean, and structurally high. The Uplift is also a preferred habitat for oil and gas, so its presence is strongly reflected in the geo-

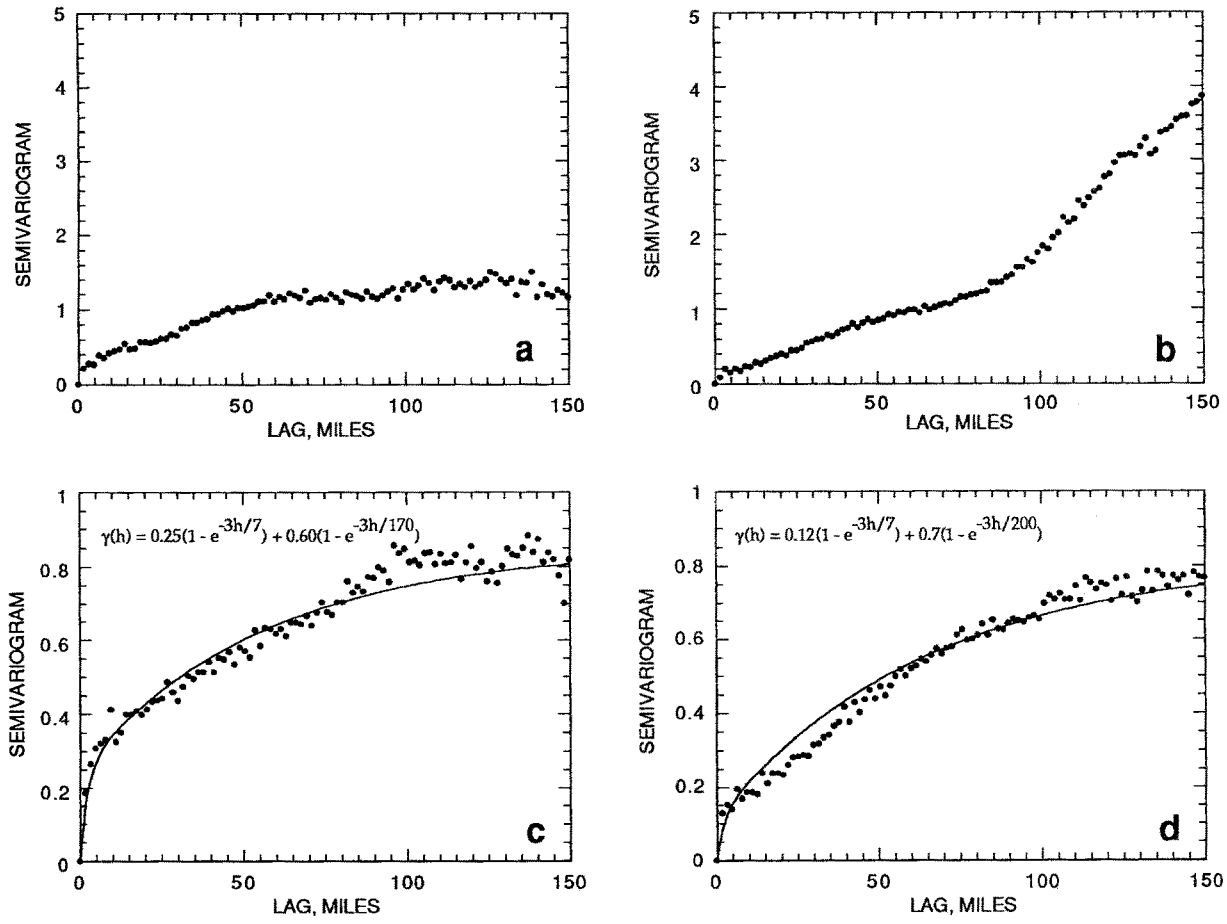
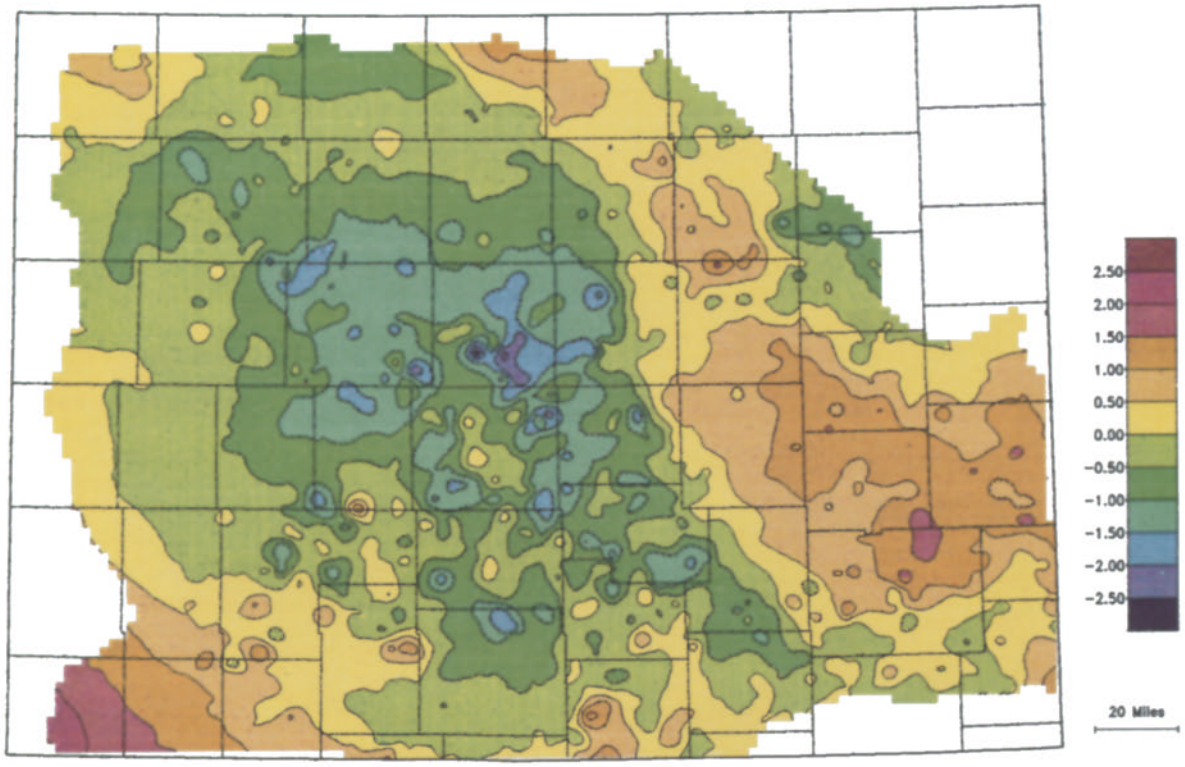


Figure 3. Variograms of Mahalanobis' distances. Both distance variables have been normalized to have means of 0.0 and standard deviations of 1.0 prior to calculating variograms. *a*, Distance to class 1, measured in NE-SW direction. *b*, Distance to class 2, measured in NE-SW direction. *c*, Distance to class 1, measured in NW-SE direction. *d*, Distance to class 2, measured in NW-SE direction. Lines in *c* and *d* are nested exponential models of the variogram. Model coefficients are given on each modeled variogram.



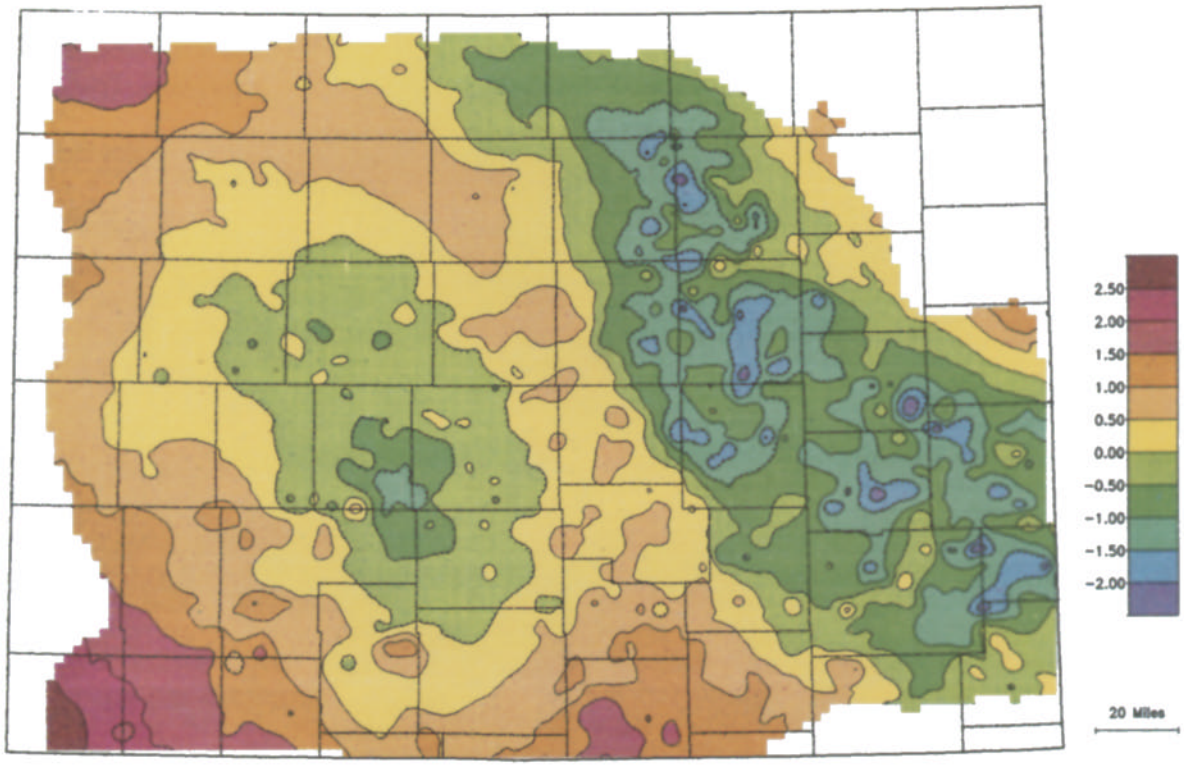
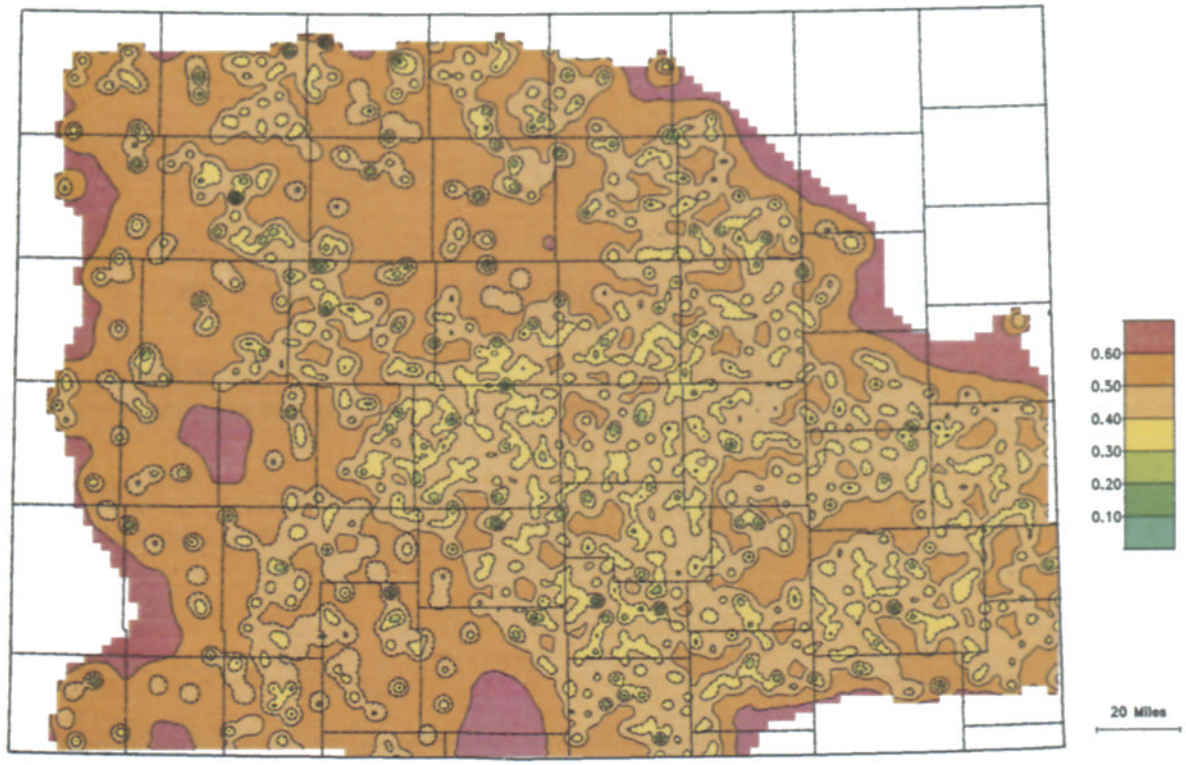


Figure 4. (Above) Contour map of Mahalanobis' distances $\hat{d}_1^{2*}(r)$ to class 1, estimated using universal kriging. Distance values have been normalized.

Figure 5. (Below) Contour map of Mahalanobis' distances $\hat{d}_2^{2*}(r)$ to class 2, estimated using universal kriging. Distance values have been normalized.



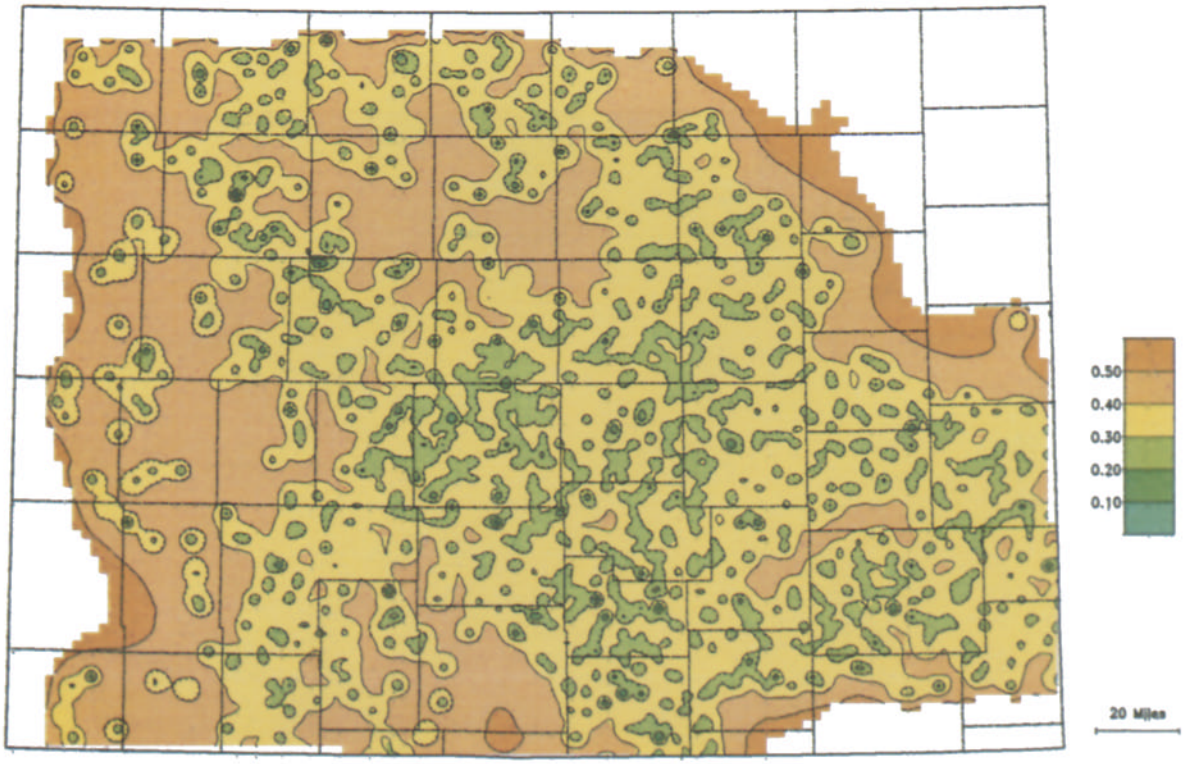
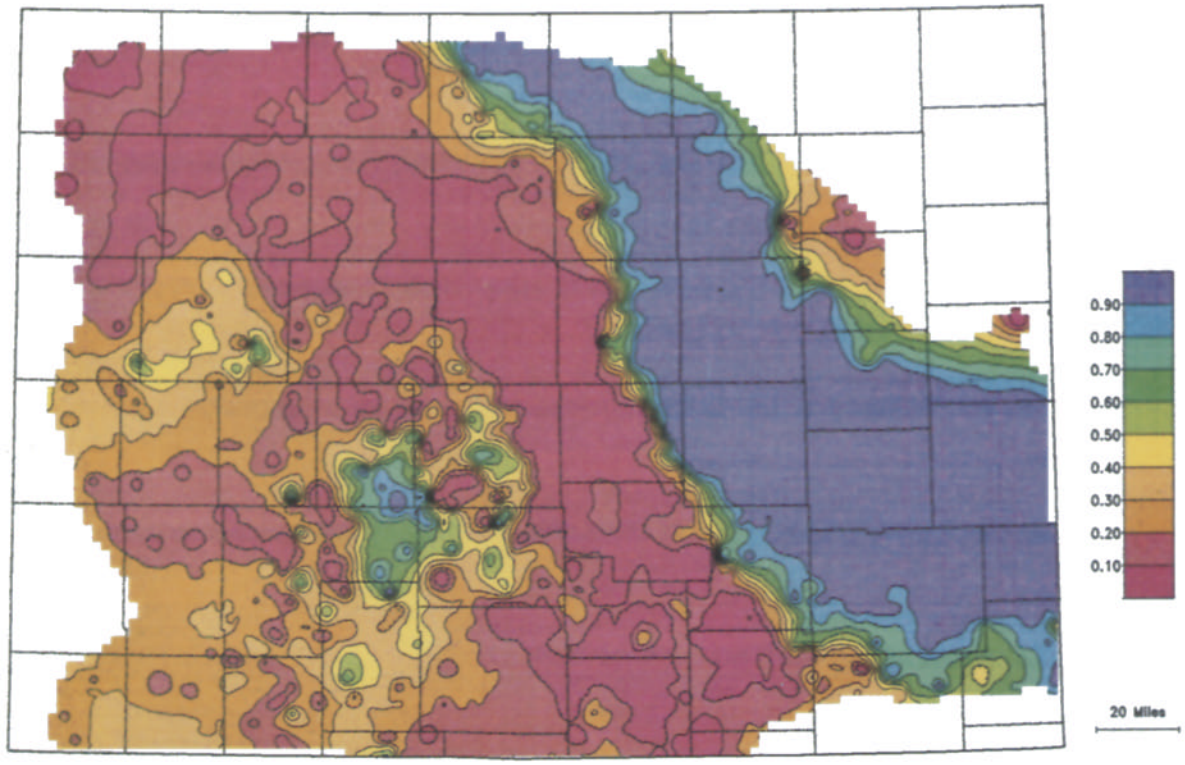


Figure 6. (Above) Kriging variance for kriged estimates of Mahalanobis' distance $\hat{d}_1^{2*}(r)$ to class 1.

Figure 7. (Below) Kriging variance for kriged estimates of Mahalanobis' distance $\hat{d}_2^{2*}(r)$ to class 2.



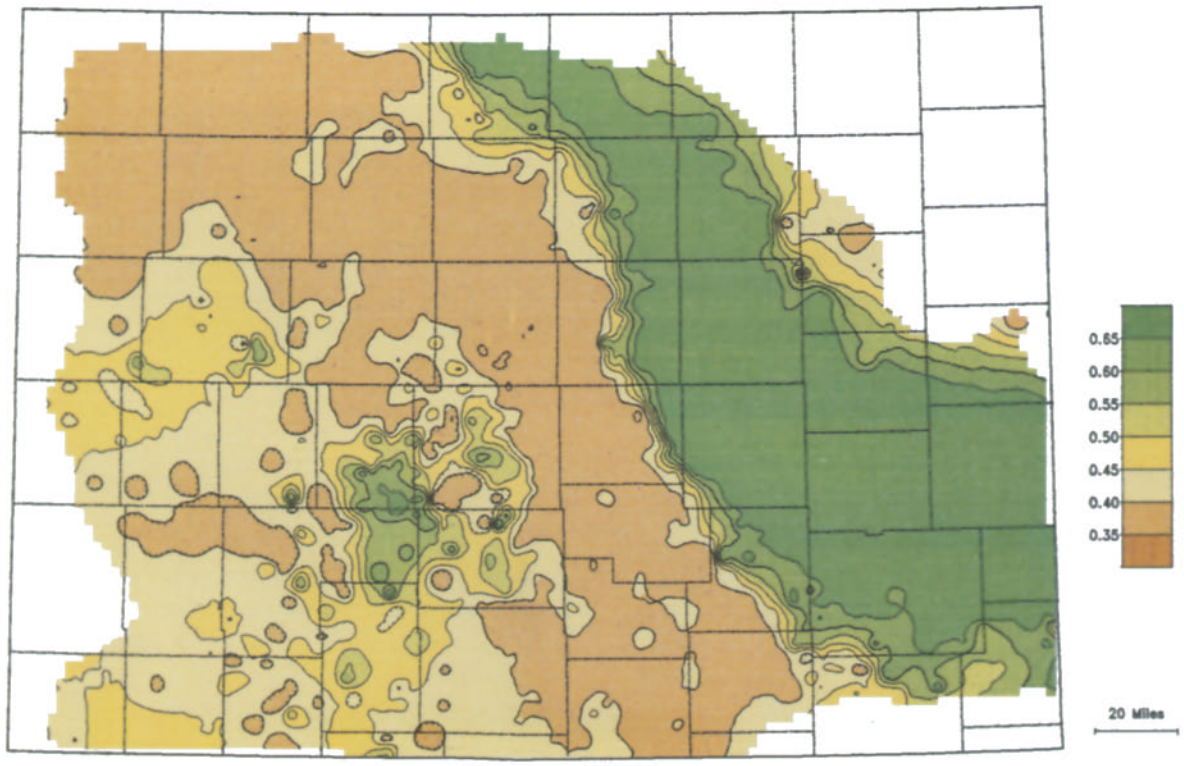


Figure 8. (Above) Probability that a well drilled in the study area will be classified as belonging to class 2.

Figure 9. (Below) Probability that drilling an exploratory hole in the study area will result in the discovery of oil.

logical characteristics of class 2, which contains most of the producing wells.

Parallel to the Central Kansas Uplift, the distance measures are drift free, as confirmed by the corresponding semivariograms. The experimental semivariograms in the northwest-southeast direction were modeled by nested exponential functions, which are shown on the plots. Because of the drift in the variables $\hat{d}_1^2(r)$ and $\hat{d}_2^2(r)$, universal kriging was used for interpolation and the local drift was modeled by a first-order polynomial.

The map grid contains over 13,000 grid nodes, and a kriging estimate has been made for each node. Figure 4 is a contour map of $\hat{d}_1^{2*}(r)$, the estimated value of the standardized Mahalanobis' distance to the centroid of dry class 1. The estimates have been produced by universal kriging of the values of $\hat{d}_1^2(b(r))$ calculated for each well location. Figure 5 is an equivalent contour map of $\hat{d}_2^{2*}(r)$, the estimated Mahalanobis' distance to the centroid of producing class 2.

At every node, the kriging estimation variance also has been calculated for class 1 and class 2, and is shown in figures 6 and 7. The values from these four map grids can be used in equation 14 to estimate the probabilities that each grid node location belongs to class 1 or to class 2. Since the probability is estimated at all grid nodes, the results also can be displayed as a contour map, as in figure 8. This map expresses the probability that an exploratory hole will be classified as belonging to B_2^g, the predictor class for producing wells.

However, class B_2^g does not consist entirely of producing wells, since the discriminant function is incapable of correctly classifying all of the input wells. By considering the misclassification ratios for dry holes classified in class B_2^g and producing wells misclassified in class B_1^g, Bayesian relationships can be used to express the probability that a location will be a producer, regardless of its classification (eq. 19). The probability that a location will produce oil is given by the contour map in figure 9.

Conclusions

Risk assessment in oil and gas exploration can be based on the statistical theories of classification of geological objects (Rodionov, 1981) and regionalized variables (Journel and Huijbregts, 1978). The method predicts the results of drilling a prospect at a location where no directly measured geological data are available. The procedure interpolates a well classification that is based on geological predictors and considers the uncertainty in the resulting classification that is a consequence of interpolation.

Results are expressed as probabilities of success in exploration. The probability of failure (exploration risk) can be calculated easily. The method is not specific to exploration for oil and gas, but also may be useful for predicting the occurrence of mineral and water resources.

References

Agterberg, F., 1989, Computer programs for mineral exploration: Science, v. 245, p. 76–81.

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.C., 1990, Regionalization in geology by multivariate classification: Mathematical Geology, v. 22, p. 573–588.

Hohn, M.E., 1988, Geostatistics and petroleum geology: New York, Van Nostrand Reinhold, 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: London, Academic, 600 p.

Matheron, G., 1970, La théorie des variables régionalisées, et ses applications: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fascicule 5, Ecole Nationale Supérieure des Mines, Paris, 211 p.

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

Rodionov, D., 1981, Statisticeskie Resenija v Geologii: Moscow, Izd. "Nedra," 231 p.

Tatsuoka, M.M., 1971, Multivariate analysis—Techniques for educational and psychological research: New York, Wiley, 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., in press, 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.F., eds., Computerized basin analysis: New York, Plenum.

Received August 6, 1991; accepted August 12, 1991.

This research was conducted during J. Harff's tenure as Visiting Research Scientist at the Kansas Geological Survey in 1991. The German Science Foundation provided travel funds. The support of both of these organizations is gratefully acknowledged.

Dr. F. Agterberg of the Geological Survey of Canada provided fruitful scientific discussions. Dr. W.L. Watney, Kansas Geological Survey, provided data for the case study. Technical assistance was given by Kansas Geological Survey staff members G. Bohling, J.C. Wong, R.J. Sampson, J.A. DeGraffenreid, and C. Cowan.