GENERALIZING GEOLOGICAL DATA FOR DEPOSIT MODELING

[1]DAVIS, J. C. and [2]HARFF, J. [1]Kansas Geological Survey, Lawrence, KS, USA; [2]Baltic Sea Research Institute Warnemünde, Germany

Summary

Regionalized classification serves as a method for data reduction in deposit modeling. The procedure combines methods of classification of geological objects with those of geostatistics. In the first step, typification, data from a random sample taken in an area of investigation are allocated to a limited number of classes represented by experimental expected value vectors and covariance matrices. In the following step, regionalization, results of typification are extended over the entire area of interest by interpolation using geostatistical methods. A two-dimensional approach is used to assess the regional favorability of hydrocarbon production. Three-dimensional regionalization can be used to produce condensed representations of the physical properties of reservoirs to be used as input for fluid flow models. Examples from the midwestern United States are presented.

Introduction

"Deposits," including ore bodies, hydrocarbon reservoirs, and aquifers, are geological bodies which can be characterized by variables measured in three-dimensional space. Interpolation and other predictive methods can be used to fill spaces between observations, resulting in exceedingly large, multidimensional numeric arrays. The enormous memory requirements for these arrays limits our ability to make dynamic simulations, even with supercomputers, and spatial and temporal resolution sometimes must be sacrificed. Generalized information for deposit assessment can be derived from very large data sets by multivariate statistical methods, simplifying the starting point for process modeling. One generalization method is "regionalized classification" (Harff and Davis, 1990), which subdivides a volume of space into regions that are as uniform as possible in their properties and can be described by a small number of statistical parameters; regionalization reduces the volume of data in a manner that preserves the natural structure of geological bodies. The reduced data model can be used in the two-dimensional case for approximate visualization of geology using GIS software to identify favorable areas for mineral exploration. Three-dimensional representations of sedimentary basins can be built up from a sequence of two-dimensional regionalization models for a limited set of sedimentary layers, which are then stacked according to information on subsurface elevations. The resulting three-dimensional characterization of a reservoir or aquifer can be used to provide input to a fluid flow model.

The model

We can represent a subunit of the Earth's crust under investigation as a random field

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

where $\mathbf{X}$ denotes an n-dimensional random variable, $\mathbf{m}$ denotes the expected value vector function, $\mathbf{Y}$ denotes the n-dimensional fluctuation, and $\mathbf{r}$ stands for the vector of coordinates within the geographic space of investigation, R.

We assume that the space R is subdivided into subunits

$$\mathbf{r}_i \in R_{ij},\ i \in I,\ j \in J_i,\ R_{ji} \subset R$$

each regarded as homogeneous in the sense that the expected value vector $\mathbf{m}(\mathbf{r}_i)$ and the covariance matrix $\mathbf{\Sigma}(r_i)$ are constant within each area R_{ij}:

$$\mathbf{m}(\mathbf{r}_i) = E[\mathbf{X}(\mathbf{r}_i)] = \mathbf{m}_i$$

and $$\mathbf{\Sigma}(r_i) = E\left[(\mathbf{X}(r_i) - \mathbf{m}_i)(\mathbf{X}(r_i) - \mathbf{m}_i)'\right] = \mathbf{\Sigma}_i .$$

Given a random variable vector, $\mathbf{X}(\mathbf{r})$, the squared Mahalanobis' metric

$$d_i^2(\mathbf{r}) = (\mathbf{X}(\mathbf{r}) - \mathbf{m}_i)' \mathbf{\Sigma}_i^{-1} (\mathbf{X}(\mathbf{r}) - \mathbf{m}_i)$$

which can be treated as a random field expresses the distance between random vector $\mathbf{X}(\mathbf{r})$ and expected value vectors $\mathbf{m}_i$.

Regionalized Classification Procedure

The procedure requires a random sample of geological bodies representing an area of investigation

$$b(\mathbf{r}_j) \in B,\ \mathbf{r}_j \in R,\ j \in N, N = \{1,...,n\}.$$

The random sample is described by a matrix of measured data $(\mathbf{x}(\mathbf{r}_j)),\ \forall j \in N$. The first step—typification—is the application of a classifier (e.g., cluster analysis) which agglomerates the elements of B, generating natural classes represented by the division $D = \{K_i\},\ i \in I^*$. Using the observed data, expected value vectors and covariance matrices for each class can be estimated, generating a type model $K_i : \{\mathbf{m}_i^*, \mathbf{S}_i\}, \forall i \in I^*$ describing the classes.

In the regionalization step, the classification (typification) of the elements of the random sample is extended to the space of investigation using an interpolator. Mahalanobis' distances $d_i^2(\mathbf{r}_j)$ can be calculated for each sampling site $j \in N$ and interpolated to nodes $\mathbf{r}_e \in R^A, R^A \subset R$ of a three-dimensional array by geostatistical methods (e.g., kriging)

$$d_i^{2*}(\mathbf{r}_e) = \sum_{j \in N} \lambda_j \, d_i^2(\mathbf{r}_j) .$$

These Mahalanobis' distances at each grid node can be transformed into class membership probabilities $p(i|\mathbf{X}(\mathbf{r}_e))$ for classes $i \in I^*$ using Bayes' theorem.

Applying the concepts of discriminant analysis at each grid (array) node $\mathbf{r}_e$, the most probable class can be chosen and marked by an element $i \in I^*$

$$D(\mathbf{r}_e) = \left\{ i \in \{I^*\} : p(i|\mathbf{X}(\mathbf{r}_e)) = \max_j p(j|\mathbf{X}(\mathbf{r}_e)) \right\}.$$

The probabilities

$$p_D(\mathbf{r}_e) = \max_{j \in I^*} p(j|\mathbf{X}(\mathbf{r}_e))$$

yield an estimate of the reliability of regionalization.

Regional Assessment of Resource Favorability

For the task of regional assessment, two-dimensional regionalization can be employed, in which case $R = R^2$. B is here a set of exploration wells, and $(\mathbf{x}(\mathbf{r}_j))$ denotes a matrix of observed values of variables useful for predicting the favorability of a desired resource. Data vectors are assigned to well sites $\mathbf{r}_j$. A classifier distinguishes between wells of different levels of favorability for oil and gas production. By interpolation the classification procedure is extended to the entire area of interest R^2. The result is a regionalization model

$$M^2 = \{D(\mathbf{r}_e), p_D^*(\mathbf{r}_e)\}, \forall r_e \in R_A^2$$

which can be visualized using GIS tools.

An example is provided by an investigation of thicknesses of genetic stratigraphic units (GSUs) of the late Carboniferous on the Kansas shelf bordering the Arkoma and Anadarko basins (Watney, *et al.*, 1999). The data were taken from a database including results from exploration and production wells. The thickness data serve as the predictor variables mentioned above. The resulting regionalization scheme divides the area of investigation into 15 basin/shelf subunits, each representing different sediment accumulation regions. The regionalization model may have significant implications for the exploration for stratigraphic traps, including the identification of migration paths along the fracture systems indicated by transition zones between the crustal blocks represented by the regions.

Three-dimensional modeling (layered model)

For three-dimensional modeling of sedimentary basin fill architecture, a hierarchical procedure is recommended. In the first step, only the linear sequences of data vectors (e.g., wire line logs) within individual wells are taken into consideration, limiting the physical coordinate space to $R = R^1$. Depth-constrained cluster analysis (Rodionov, 1981; Gill, *et al.*, 1993) of a master well and a subsequent lithostratigraphic correlation into neighboring wells (Olea, 1994) provide the means to subdivide the sedimentary sequences penetrated by the wells into zones $b(\mathbf{r}_i, r_t) \in B, \mathbf{r}_i \in R^2, r_t \in R^1, i \in N, t \in T$. Here, r_t marks the top of a zone. Each zone is described by "blocked" data (averaged within a layer) stored in the matrix $(\mathbf{x}(\mathbf{r}_j, r_t))$. The correlative zones within each well represent layers which can be treated by two-dimensional regionalization. That means that for each layer, t, a regionalization model, M_t^2, can be combined with a digital elevation model of the subsurface, created by geostatistical interpolation procedures,

$$DEM_t = \left\{ r_{te} \in R^2 : r_{te} = \sum_j \lambda_j \, r_{tj} \right\},$$

to create a three-dimensional data model,

$$M^3 = \{M_t^2, DEM_t\}_{t \in T}.$$

As a case study the Zenith oil field in central Kansas provided a convenient database for testing regionalization procedures (Harff *et al.*, 1991). The production comes from five different sandstone and carbonate reservoirs which have been typified separately by cluster analysis using three wireline-log variables: gamma-ray intensity, corrected neutron porosity, and corrected density porosity. Five rock types for the carbonates and four rock types for the sandstones were identified. The stratigraphic zones were correlated using the program CORRELATOR (Olea, 1994), based on gamma ray and neutron porosity logs from 38 wells. Within each stratigraphic zone, wells were assigned to rock types (classes) by discriminant analysis and the probabilities of class membership were interpolated to the nodes of a regular grid. This procedure subdivides each layer into homogeneous regions characterized by class mean values of the petrophysical variables. Additional variables, such as porosity, were computed using deterministic models based on wireline-log responses and were then averaged within classes. Permeabilities were calculated from regression between core permeabilities and wireline-log responses for the equivalent intervals. The regression was used to estimate a class mean permeability from the means of gamma-ray intensity and porosity. The data model, consisting of a stack of grids, can be displayed graphically on a workstation and serves as a basis for further investigation using a black-oil reservoir simulator.

References

Harff, J. and Davis, J.C., 1991. Regionalization in Geology by Multivariate Classification. Mathematical Geology, v. 22, no. 5: 573–588.

Harff, J., Davis, J.C. and Eiserbeck, W., 1993. Prediction of Hydrocarbons in Sedimentary Basins. Mathematical Geology, v. 25, no. 7: 925–936.

(cont.)

Harff, J., Davis, J.C., Olea, R.A., Watney, W.L., Doveton, J.H., Bohling, G.C., Newell, K.D., McElwee, C.D., Hoth, P., Lewerenz, B., Springer, J., Sampson, R.J., Wong, J.C. and Cunnigham, K.J., 1991. Three-Dimensional Regionalization and Modeling for Sedimentary Basin Analysis: The Zenith Oil Field, I. Kansas Geological Survey OFR 91–41: 95, Lawrence.

Gill, D., Shomrony, A., and Fligelman, H., 1993. Numerical Zonation of Log Suites and Log Facies Recognition by Multivariate Clustering. AAPG Bulletin, v. 77, no. 10: 1781–1791.

Olea, R.A., 1994, Expert System for Automated Correlation and Interpretation of Wireline Logs. Mathematical Geology, v. 26, no. 8: 879–897.

Watney, L.W., Kruger, J.C., Davis, J.C., Harff, J., Olea, R.A. and Bohling, G.C., 1999, Validation of Sediment Accumulation Regions in Kansas, USA. In: Harff, J., Lemke, W. and Stattegger, K. (Eds.) Computerized Modeling of Sedimentary Systems. Springer–Verlag: 341–360, Berlin.