in Current Trends
in Geomathematics
D.F. Merriam (ed)
Plenum Press 1988

KRIGING HYDROCHEMICAL DATA

Donald E. Myers

University of Arizona

ABSTRACT

As a part of the National Uranium Resource Evaluation Program (NURE) water samples were collected from existing wells in all the continental United States. These samples were analyzed for some 30 elements and ions. Data were assembled for each 2 degrees RMTS quadrangle. The objectives of the NURE program included identification of areas favorable for exploration and producing estimates of recoverable resources. Other authors have reported on the use of pattern recognition, cluster analysis, and discriminant analysis to identify favorable areas.

In cooperation with the Uranium Resource Evaluation Group at Oakridge, the author utilized data from Plainview Quadrangle (Plainview, Texas) to examine the effectiveness of kriging to contour data on 13 variables including uranium. These variables were selected for their chemical association with the deposition or leaching of uranium salts. Because of strong dissimilarities between the Ogallala (Pliocene) and Permian groupings, the data were segregated.

Variograms were computed for each variable, separately for the Permian and Ogallala. Variogram models were cross-validated using randomly selected data subsets. In addition to kriged contour maps for the 13 variables and kriging variance maps in

both the Permian and Ogallala, weighted linear sums also were considered. Two different weightings were considered, the weights were determined by a discriminant analysis model. Unusual regions were identified as those for which the kriging error exceeded two kriging standard deviations. These regions were correlated strongly with those identified by a discriminant analysis model and by the quadrangle evaluation.

INTRODUCTION

The objective of the National Uranium Resource Evaluation (NURE) Program was "to provide a systematic appraisal of the uranium resources of the conterminous United States and Alaska" Everhart (1977). It was envisioned that geologic, radiometric, hydrogeochemical, and stream-sediment data would be collected and analyzed systemically in an appropriate manner and evaluations prepared for various geographical regions. The Hydrogeochemical and Stream Sediment Reconnaissance (HSSR) Program was one facet of the data collection process. As indicated by Roach (1978) it was expected that statistical analysis would play an important role but the types of techniques to be used were not specified. Kane (1977) has described the application of standard statistical techniques such as cluster and factor analysis to HSSR data. This paper will present the results of applying kriging to hydrogeochemical data from the Plainview (Texas) Quadrangle (NTMS).

Kriging is a linear estimation technique that incorporates the spatial dependence of the variable in question. Kriging was investigated as a tool to delineate geochemical patterns, identify anomalous areas, and dispersion properties of hydrogeochemical variables in a quantitative way.

THE PLAINVIEW QUADRANGLE AND HSSR DATA

As a part of the HSSR Program, water samples were obtained from approximately 900 wells in the Plainview Quadrangle which then were analyzed at the ORGD analytical laboratory. Each sample location was identified by latitude and longitude and observed values recorded for some thirty hydrogeochemical variables. A complete listing is given in the open-file quadrangle report (URE, 1978) and the data also are available.

A detailed discussion of the geology of the Plainview Quadrangle and the reasons for its selection for this study is contained in Myers and others (1980). Briefly, the reasons include the following: (1) relative simplicity of the geology, there being only two major geologic formations; (2) the Plainview Quadrangle was the only one for which the quadrangle evaluation was complete (Amaral, 1979); and (3) good overall groundwater sample coverage. There are 473 sites in the Permian units and 375 in the Ogallala Formation. Data for twelve variables were considered in the Permian and thirteen in the Ogallala. These will be listed later.

KRIGING

The statistical technique known as kriging was developed by Matheron (1965, 1971, 1973) and his associates at the Centre de Geostatistique, ENSMP, France to provide an improved method of ore-grade estimation. It also has been used as a contouring technique in hydrology and more recently for soil mapping (Journel and Huijbrechts, 1978; Burgess and Webster, 1980). The application to hydrogeochemical data reported here apparently is new and utilizes kriging for more than just contouring.

The reader is referred to Journel and Huijbrechts (1978) or Myers and others (1980) for a more complete derivation of the kriging estimator and its properties; the following is a brief summary.

Let x be a geographical position and $z(x)$ the value of a hydrogeochemical variable such as uranium concentration at x. $z(x)$ can be considered as a function defined on a two-dimensional region, well and aquifer depths were not used. If the form of the function were known it would be sufficient to substitute simply the coordinates for x and compute $z(x)$. $z(x)$ is in general an irregular function and its form is not known; only the values $z(x_1), \ldots, z(x_n)$ at sample locations $x_1, x_2, \ldots, x_n$. The problem then is to estimate or predict the value at an unsampled location. Inverse Distance Weighing (IDW) and Polygonal are two widely used methods both of which incorporate local influences. The parallel study of the Plainview Quadrangle data using IDW is reported in Kane and others (1982) and is also in Myers and others (1980). Trend-Surface Analysis (TSA) attempts to fit a smooth function to the data and does not incorporate local influences. To derive the kriging estimator it is assumed that z(x) is a realization of a random function Z(x). It then is necessary to determine appropriate statistical characteristics of Z(x) to proceed with estimation. Matheron determined that two conditions were sufficient

$$E\,[Z(x) - Z(x+h)] = 0 \qquad (1)$$

for all x, h (h a vector)

$$\mathrm{Var}\,[Z(x) - Z(x+h)] = 2\gamma(h) \qquad (2)$$

where $\gamma(h)$ depends only on h. $\gamma(h)$ quantifies the spatial dependence. Equation (1) implies the absence of drift. If $\gamma(h) = \gamma(|h|)$, $|h|$ = length of h, Z(x) is said to be isotropic. The kriging estimator is of the form

$$Z^{*}(x) = \sum_{j=1}^{n} \lambda_j(x)\; Z(x_j) \qquad (3)$$

where the λ_j 's are selected so that Z^* is an unbiased estimator, that is

$$E[Z^{*}(x) - Z(x)] = 0 \qquad (4)$$

and the variance of the error is minimal

$$\mathrm{Var}\,[Z^{*}(x) - Z(x)] = \sigma_K^2\,(x) \qquad (5)$$

The minimal value σ_K^2, is termed the kriging variance. The λ_j's are obtained from the linear system

$$\begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} & 1 \\ & \cdot & & \\ & \cdot & & \\ \gamma_{n1} & \cdots & \gamma_{nn} & 1 \\ 1 & & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \cdot \\ \cdot \\ \gamma_n \\ 1 \end{bmatrix} \qquad (6)$$

where $\gamma_{ij} = \gamma(x_i - x_j)$, $\gamma_i = \gamma(x - x_i)$ and μ is a Lagrange multiplier introduced to solve the minimization problem. To apply kriging it is necessary to test whether conditions (1) and (2) are satisfied and to determine $\gamma(h)$ which is termed the variogram. $\gamma(h)$ can be estimated by the sample variogram

$$\gamma^*(h) = \frac{1}{2N} \sum (z(x+h) - z(x))^2 \qquad (7)$$

where N is the number of pairs of sample locations at "distance" h and the sum is over all such pairs. It is known that $-\gamma(h)$ must be conditionally positive definite and the usual procedure is to try to fit $\gamma^*(h)$ to one of several known standard functional types. For ore-grade estimation there is a moderate amount of accumulated experience which provides guidance on selecting a functional form for $\gamma(h)$. For example, if a spherical model is used for $\gamma(h)$, then the parameters are related in a direct way to the graph of $\gamma^*(h)$. Because this was a new application there are no references to previous studies. It was determined that the sample variograms also provided insight into the hydrogeochemical groupings.

SAMPLE VARIOGRAMS

Because the HSSR data were collected for the purpose of aiding in the assessment of uranium resources the principal variable of interest was uranium, the other variables were selected because of their usefulness in identifying or predicting uranium occurrences. Sample variograms (svg) were computed and plotted for the following elements or variables: Uranium, Boron, Barium, Calcium, Lithium, Magnesium, Molybdenum, Sodium, Vanadium, Sulfate, Specific Conductance, Total Alkalinity, and Arsenic (Ogallala only). Because of the small number of sample locations in the Dockum Group these were omitted. The svg's were computed and plotted separately for the Permian and the Ogallala. To test whether an isotropic model for γ could be used, directional plots also were made. Because the sample locations were not on a uniform grid few pairs had the same distance although the total number of pairs is large. For plotting purposes, N was taken to be 1000 and the plotted value was an average. Because it is general practice to fit geochemical data to a log normal distribution svg's also were computed and plotted for logarithmic transformed data.

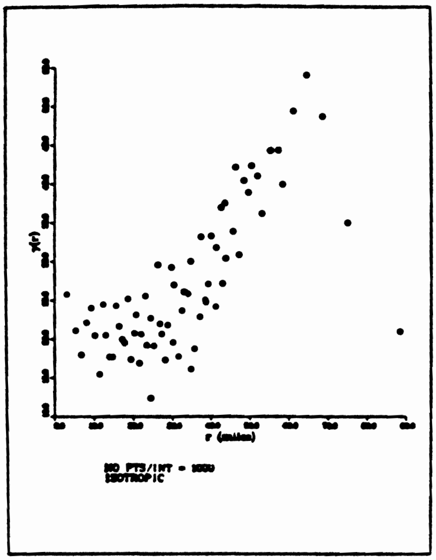
Figure 1 shows the plots for uranium, directional and isotropic for the Ogallala Formation. Figure 2 shows the same plots for the Permian. Figure 3 illustrates the possible contrast between the Ogallala and Permian units. Figure 4 illustrates how the svg differentiates between variables.

As described in Myers and others (1983) the variograms were of four graphical types.

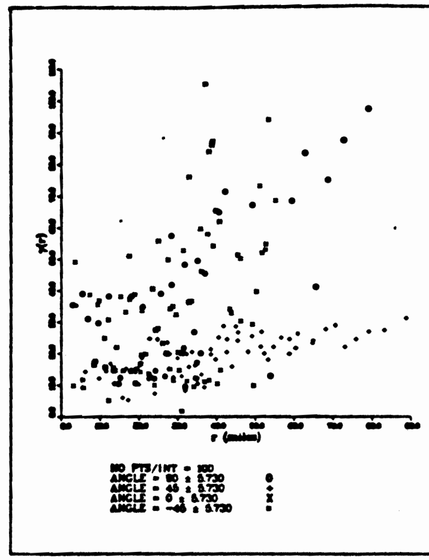
It is of interest to note that all of the plots exhibited a "nugget effect". A complete set of the svg plots and a description of the computer program is presented in Myers and others (1980)

FITTING THE SEMIVARIOGRAMS

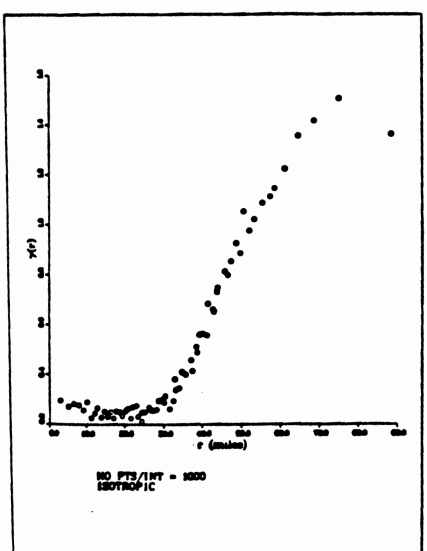
After computing and plotting the svg's these plots were used to determine whether anisotropic models must be used, the type of functional forms that might be appropriate, and the range of influence. A subjective decision was made to use only isotropic models and the functional forms to be used were determined to be power, logarithmic, and exponential. It was assumed that the range of dependency would be considerably less than 90 miles and in fact functions were fitted only on the first 30 miles. A general characteristic of the kriging estimator that the coefficients, for locations far away, will be small. As yet there no adequate statistical tests for determining the best model. Having identified an appropriate functional form for each



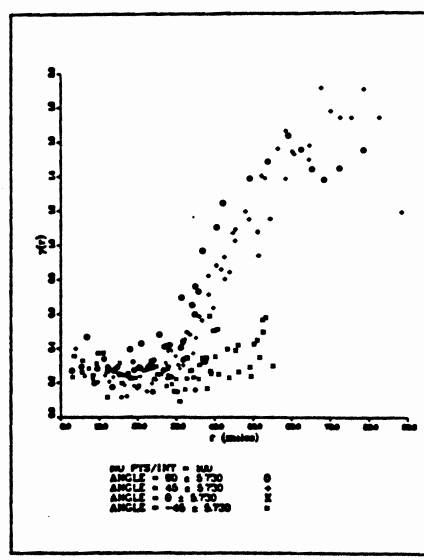
a. Isotropic - Untransformed Data



b. Anistropic - Untransformed Data

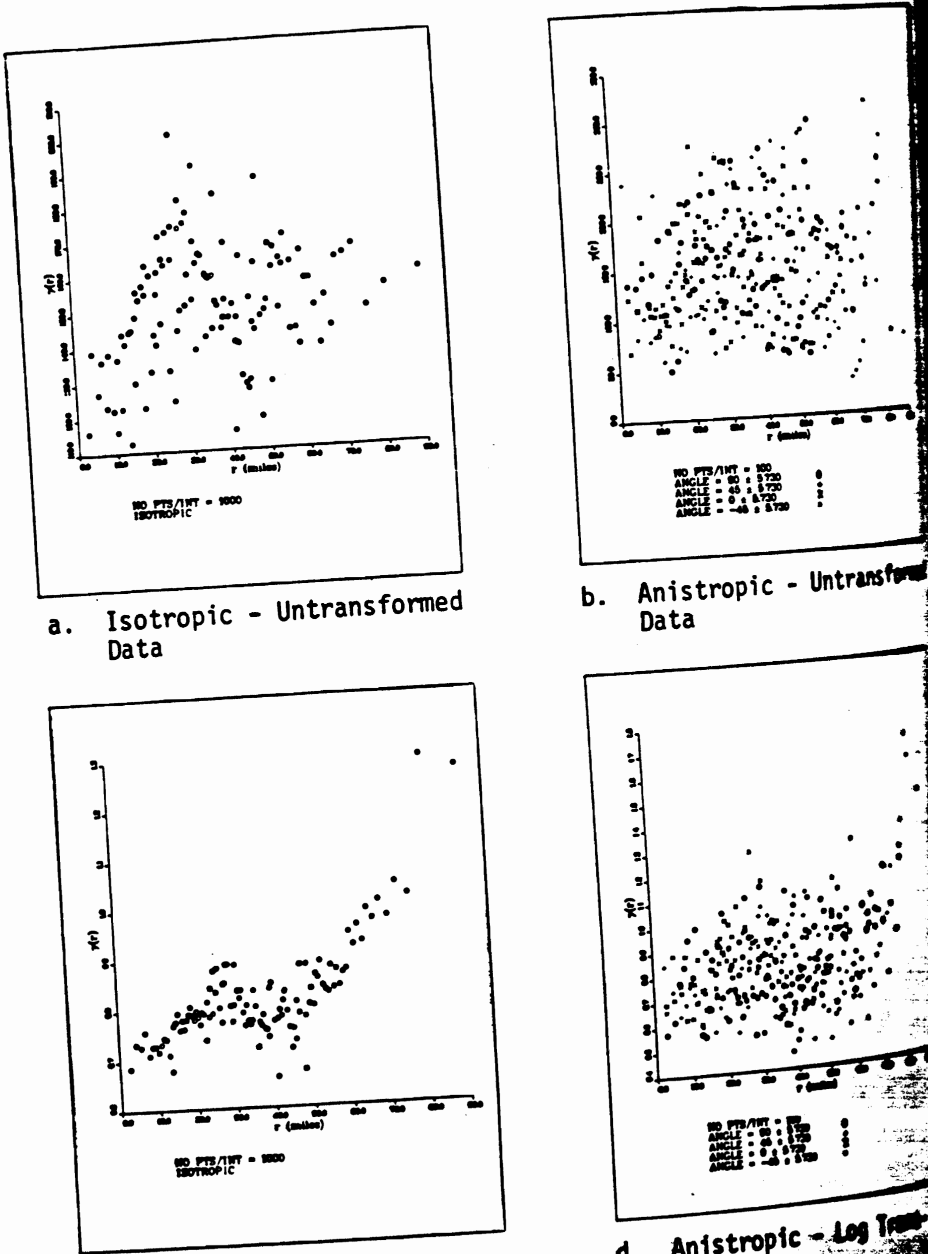


c. Isotropic - Log Transformed Data



d. Anistropic - Log Transformed Data

Figure 1. Uranium semivariograms for Ogallala Formation.



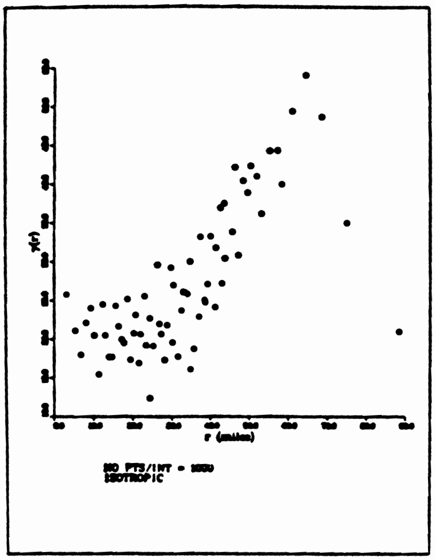
a. Isotropic - Untransformed Data

b. Anistropic - Untransformed Data

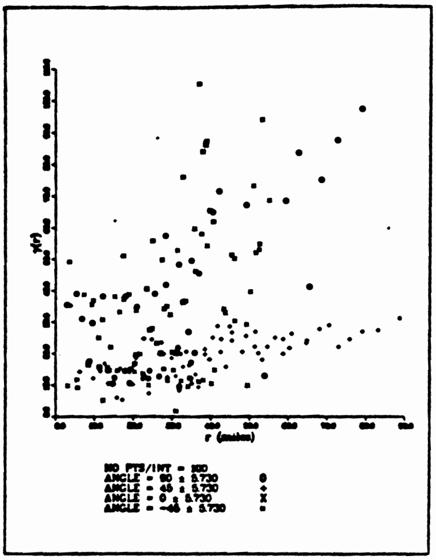
c. Isotropic - Log Transformed Data

d. Anistropic - Log Transformed Data

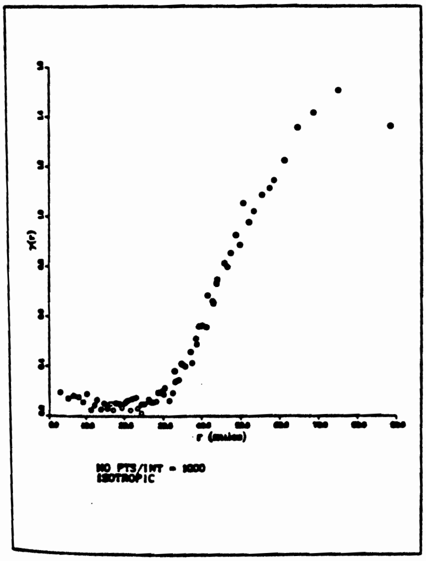
Figure 2. Uranium semivariograms for Permian units.



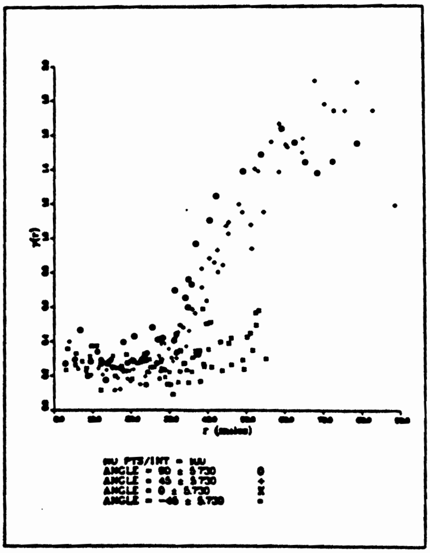
a. Isotropic - Untransformed Data



b. Anistropic - Untransformed Data

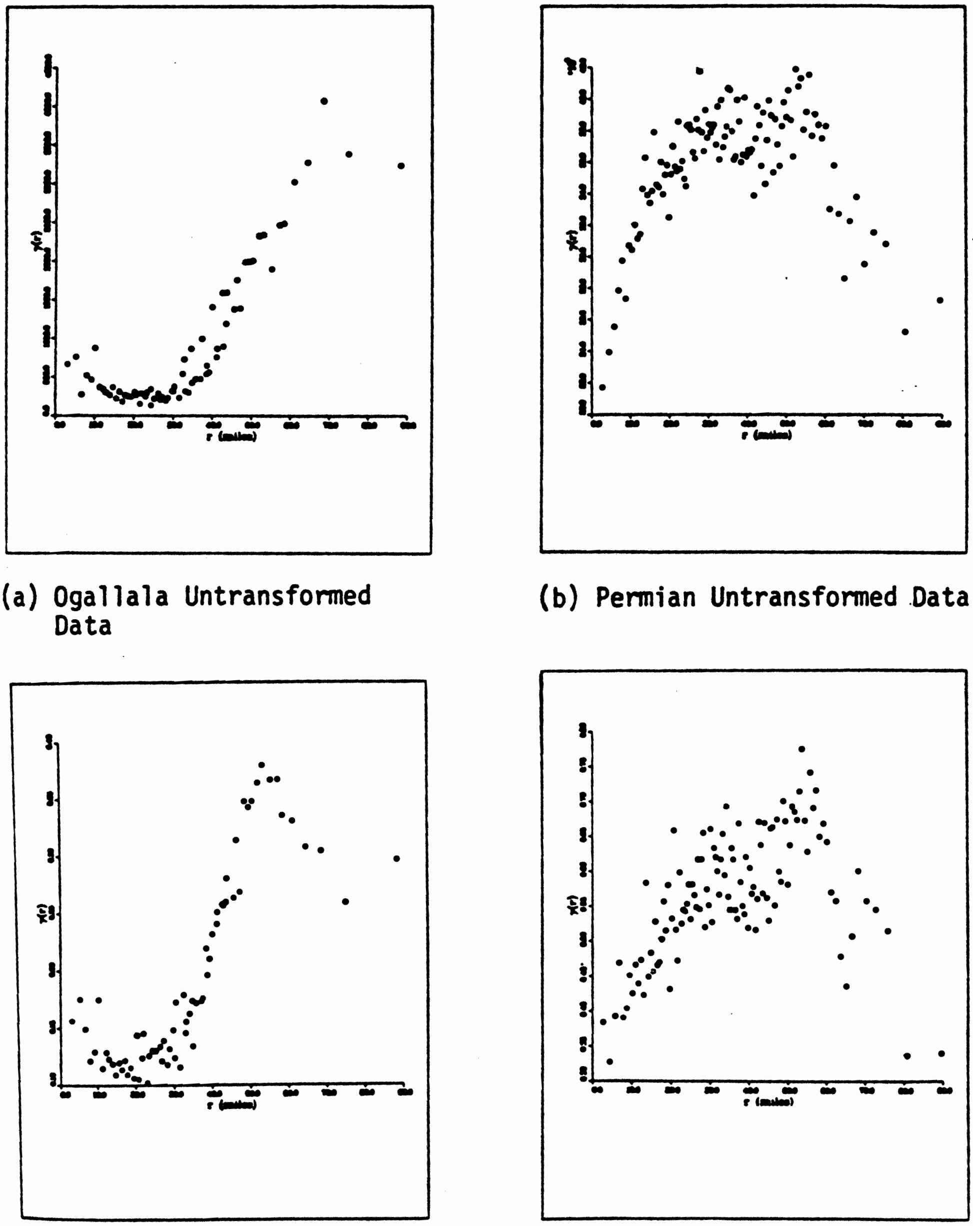


c. Isotropic - Log Transformed Data



d. Anistropic - Log Transformed Data

Figure 1. Uranium semivariograms for Ogallala Formation.



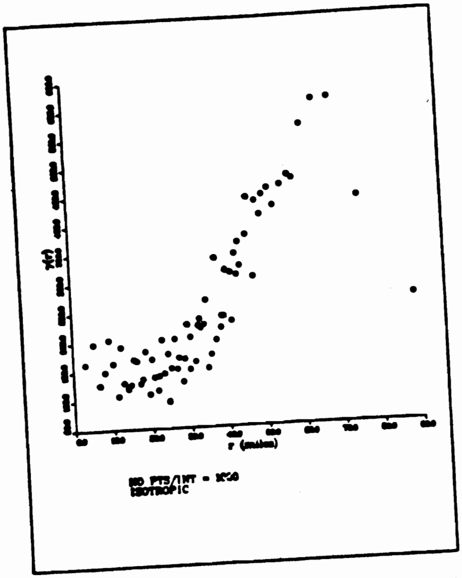
(a) Ogallala Untransformed Data

(b) Permian Untransformed Data

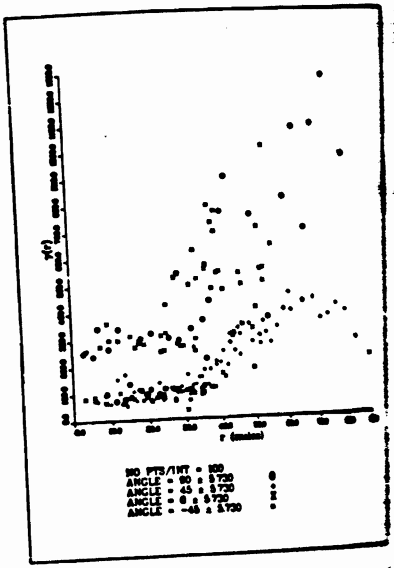
(c) Ogallala Log Transformed Data

(d) Permian Log Transformed Data

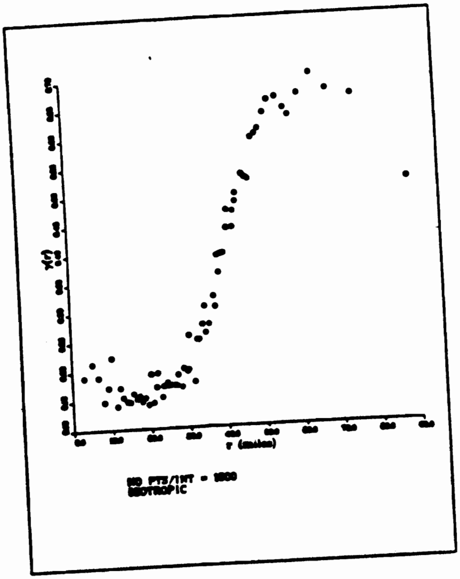
Figure 3. Calcium isotropic semivariograms for Ogallala Formation and Permian units.



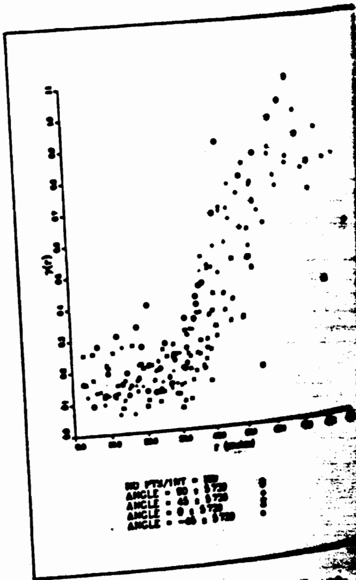
a. Isotropic - Untransformed Data



b. Anistropic - Untransformed Data

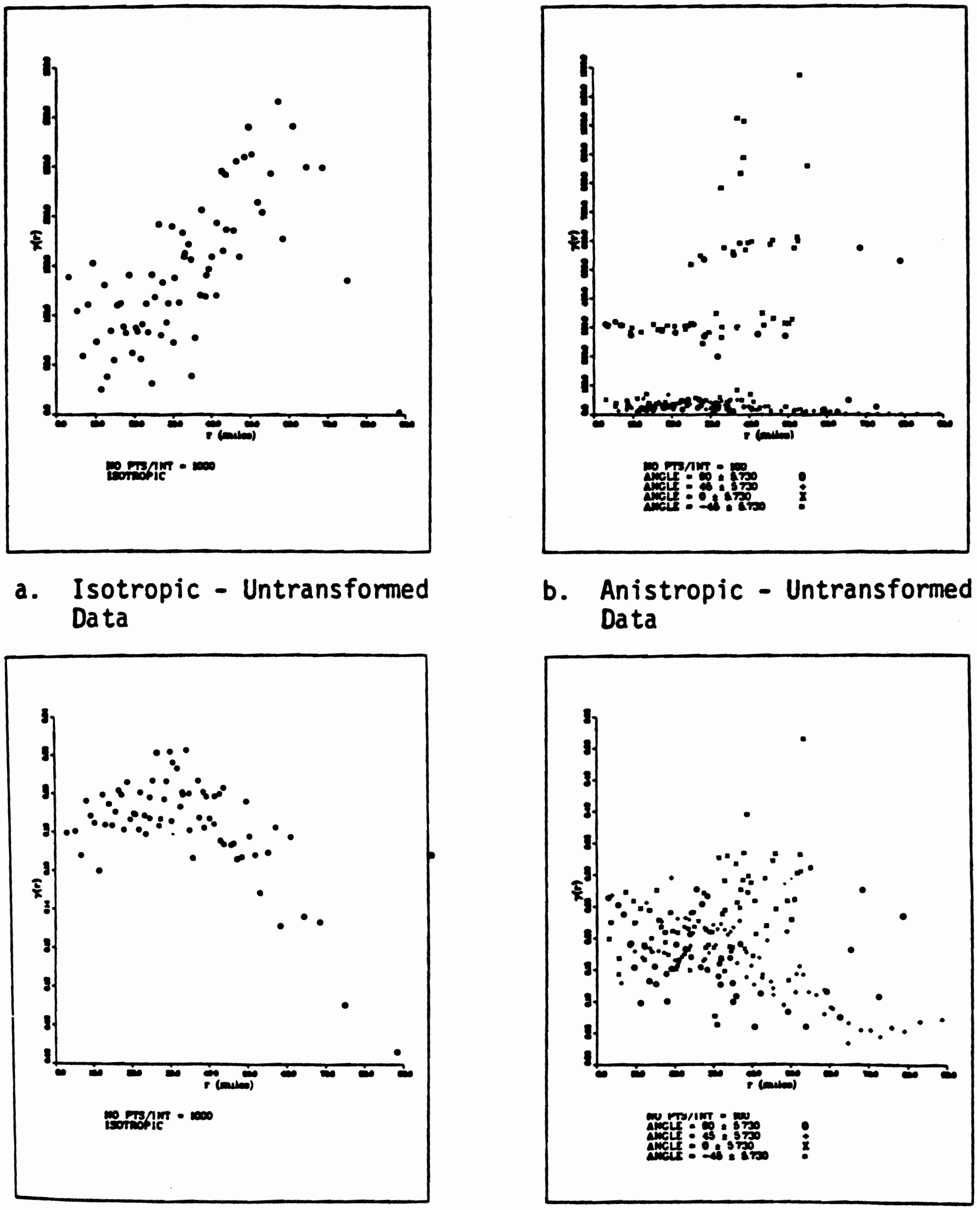


c. Isotropic - Log Transformed Data



d. Anistropic - Log Transformed Data

Figure 4-1. Magnesium semivariograms for Ogallala Formation



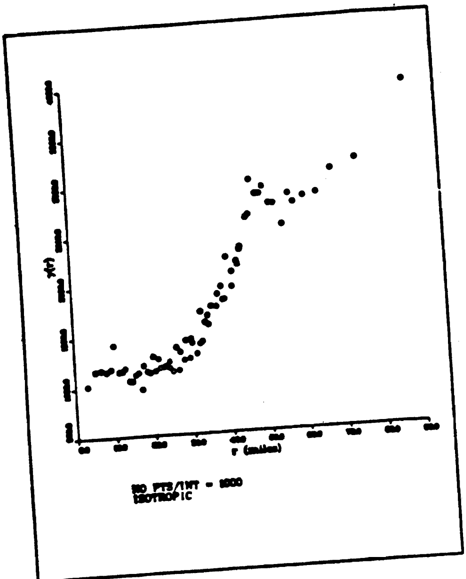
a. Isotropic - Untransformed Data

b. Anistropic - Untransformed Data

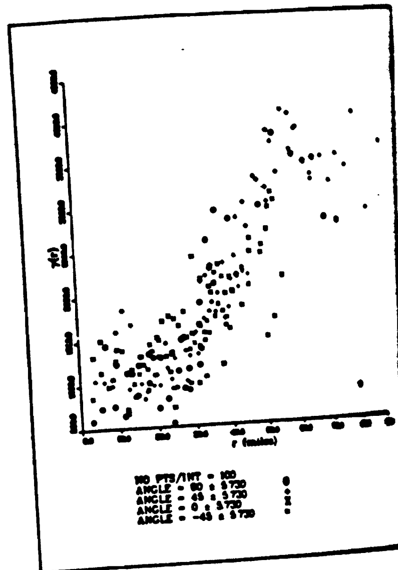
c. Isotropic - Log Transformed Data

d. Anistropic - Log Transformed Data

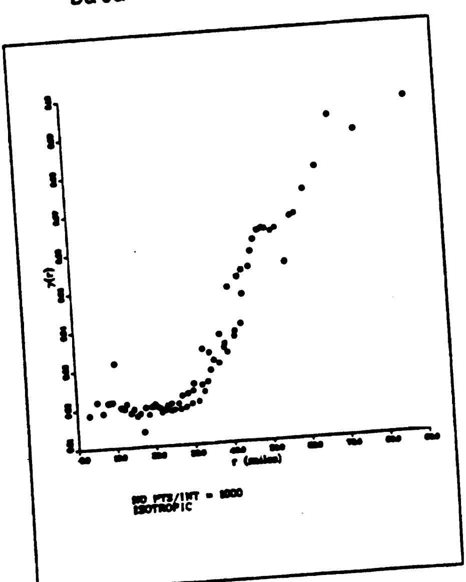
Figure 4-2. Molybdenum semivariograms for Ogallala Formation.



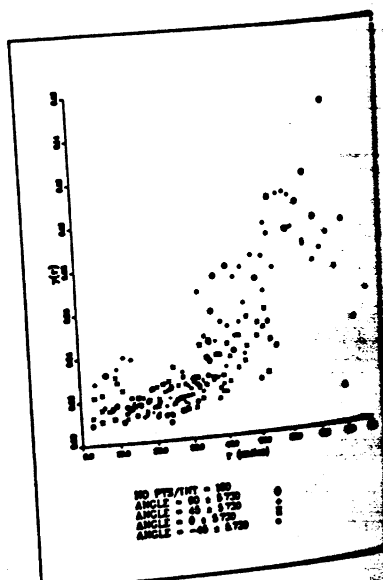
a. Isotropic - Untransformed Data



b. Anistropic - Untransformed Data



c. Isotropic - Log Transformed Data



d. Anistropic - Log Transformed Data

Figure 4-3. Total alkalinity semivariograms for Ogallala Formation and in each formation, determination of exponents or coefficients was made by standard nonlinear regression analysis. This is not necessarily defensible mathematically but does provide a practical way to fit functions to the plots. For some variables more than one form was considered and for most variables both the transformed and untransformed data were used.

Because kriging is an exact interpolation technique, an estimate can only be obtained for a sample location by deleting that location from the data set. The difference between the observed and the estimated values is indicative not only of how well the selected function form fits but also how plausible are the mathematical assumptions, of isotropy, stationarity, and zero drift. These residuals normalized by the kriging standard deviation can be used to identify unusual locations.

To compute these residuals for all sample locations and all proposed models would have required large amounts of computer time. In order to select the final functional forms to be used, only one for each variable, a random subsample of 100 locations in the Ogallala and 100 in the Permian were selected. For each of these a kriged estimate was obtained using only other sample locations within 10 miles. This was done for each proposed functional model for each variable.

A good fit for the semivariogram should produce collectively small residuals, except that large normalized residuals might indicate anomalies.

Three statistics were computed for each functional model. These were

$$S = \frac{\sum_{i=1}^{100} (z_i - z_i^*)^2}{\sum_{i=1}^{100} (z_i - \bar{z})^2} \qquad (8)$$

Table 1. S, S*, $\bar{S}$ Values

	S	S*	$\bar{S}$	Unit
U	2.033	0.105	1.330	Ogallala
L-U	0.933	1.354	0.888	Permian

$$S^* = \frac{\sum_{i=1}^{100} (z_i - z_i^*)^2 / \sigma_i^2}{\sum_{i=1}^{100} (z_i - \bar{z})^2 / s_i^2} \quad (9)$$

$$\bar{S} = \frac{\sum_{i \varepsilon D} (z_i - z_i^*)^2}{\sum_{i \varepsilon D} (z_i - \bar{z})^2} \quad (10)$$

z_i represents the observed value at location x_i, z_i^* the kriged estimated and z the sample mean. σ_i^2 is the estimation variance for location x_i and D= $\{i \,/\, |z_i - z_i^*| \,/\, \sigma_i \leq 2\}$. The denominators are almost the sample variance. S, S*, $\bar{S}$ provide a comparison between the kriging estimator and the sample mean as an estimator. Table 1 tabulates the values for S, S*, $\bar{S}$ for Uranium (Ogallala) and Log-Uranium (Permian) for the models that were used subsequently (for the 100 test locations).

The use of transformed data does introduce a bias, that is when a nonlinear transformation such as the logarithm is used the kriged estimator is nonlinear and in general not an unbiased estimator. Journel and Huijbrechts (1978) suggest ways to remove this bias but this was not incorporated in the preliminary study.

INTERPRETATION OF PLAINVIEW DATA

One of the ways that kriging can be used to interpret geochemical data is to produce contour plots. This was done for all 13 variables in the Ogallala and the 12 in the Permian. Those for uranium have been combined into one plot as shown in Figure 5. By overlaying these plots on the plot of favorable areas as determined by Amaral (1979), it is seen that there is strong coincidence of high concentration contours with areas A and B in the Ogallala.

There is some coincidence with area D in the Permian but it seems that area D should be extended southwest. In the NURE quadrangle report correlations are tabulated for each pair of variables. Those showing the highest correlation with uranium, magnesium, vanadium, lithium, total alkilinity. and arsenic (Ogallala only) do not exhibit similar contour patterns as uranium. Of perhaps equal interest are the dissimilarities between the Ogallala and Permian as exhibited in the contour plots.

The residuals described earlier also can be used to identify unusual patterns. An observed value will be termed unusual if the normalized residual is large, for example, in absolute value greater than two. The term unusual is used in contrast to anomalous because the residuals can be positive or negative. In particular for hydrogeochemical data large negative residuals may to correspond to locations where precipitation from the groundwater is taking place and may be as significant as large positive residuals. The normalized residuals have been coded onto Figure 5 and particularly in the Ogallala, exhibit a pattern which correlates with Areas A and B.

Identifying unusual values by the size of normalized residual is justified for several reasons. In Figure 6, the histograms of the normalized residuals, it is seen that the empirical probability of large residuals is small. If the normalized residuals were distributed normally then the probability of large residuals could be obtained from a normal table. If the residuals are assumed symmetric, then Chebyshev's Inequality asserts that the probability of residuals greater than 2 is less than 0.11. The symmetry that is exhibited in Figure 6 also is indicative of the unbiasedness that should be characteristic of the kriging estimator.

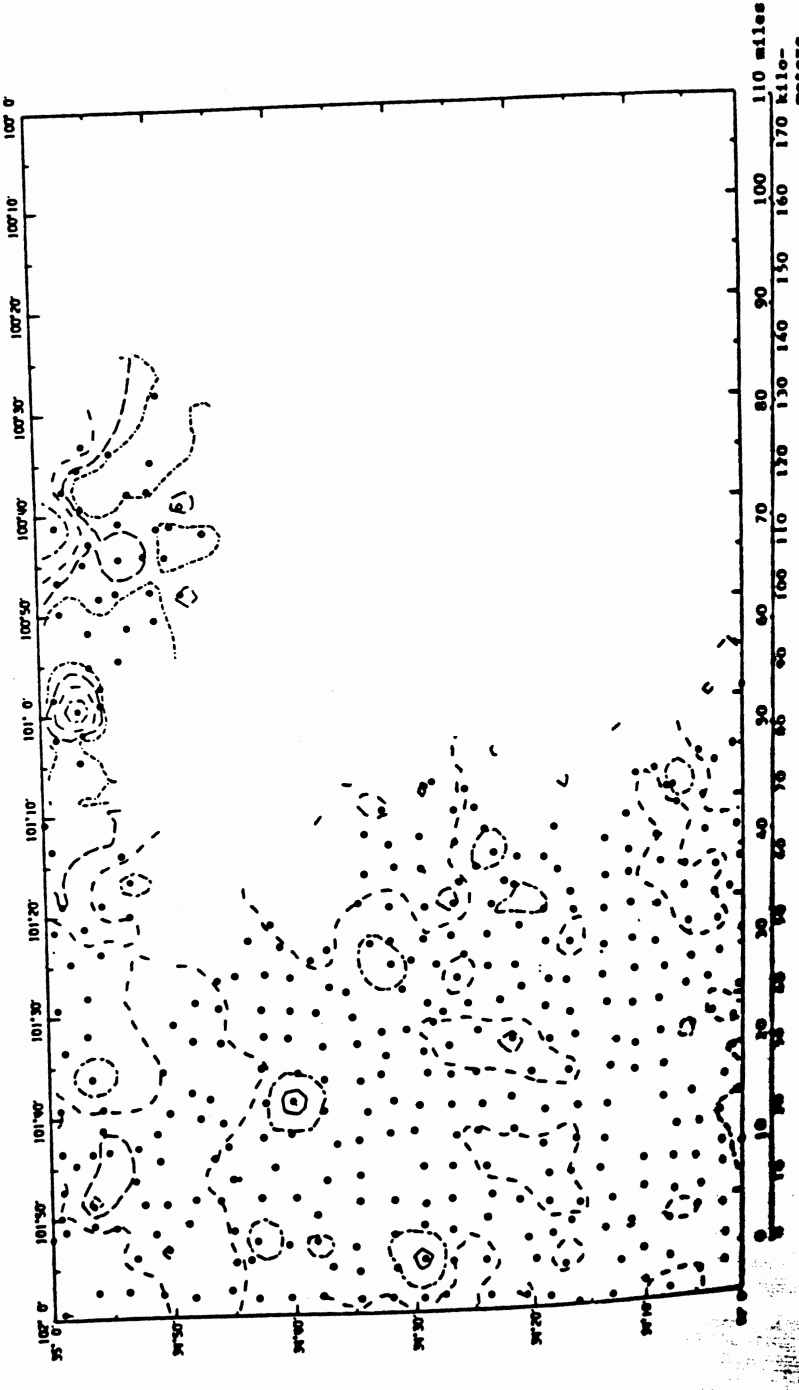


Figure B-1. Geostatistical map for uranium: Ogallala kriged.

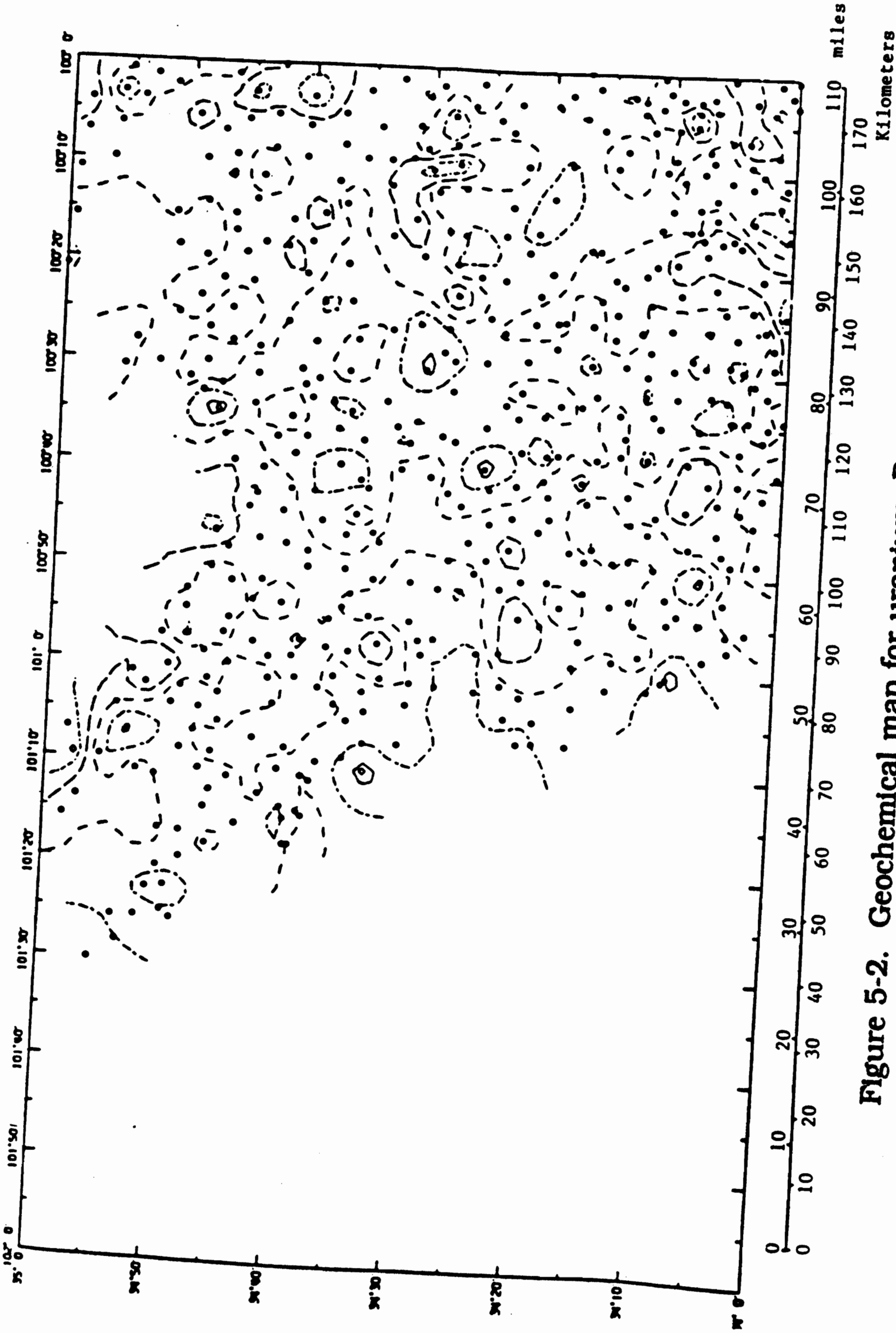
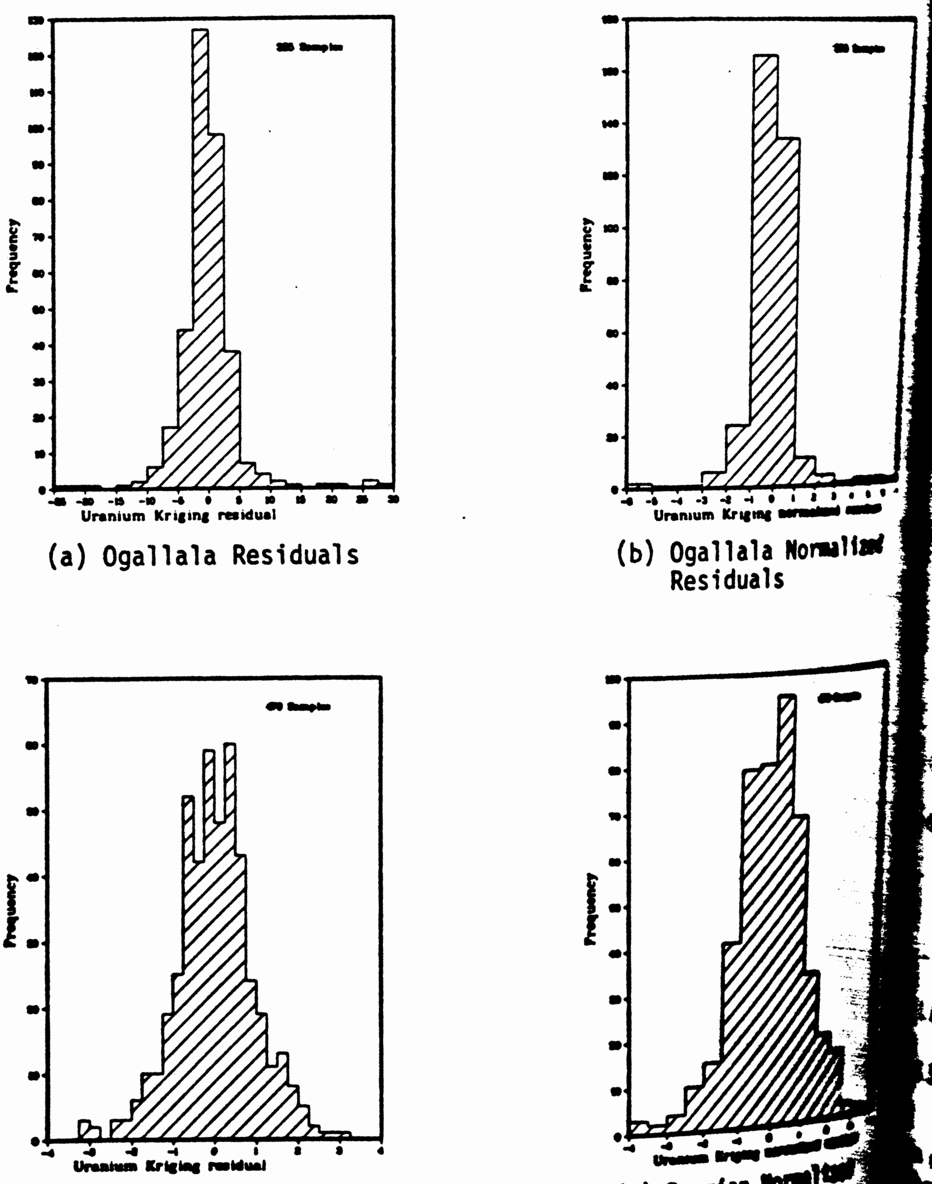


Figure 5-2. Geochemical map for uranium; Permian kriged.



(a) Ogallala Residuals

(b) Ogallala Normalized Residuals

(c) Permian Residuals

(d) Permian Normalized Residuals

Figure 6. Uranium residual histograms from kriging.

The original motivation for studying the variables other than uranium was because of their usefulness in predicting uranium mineralization. However the kriging described does not explicitly incorporate such information. There is a form of joint estimation known as cokriging which was described by Myers (1982, 1983). A somewhat simpler approach was used instead in this preliminary work. Kane (1978) has described the use of weighted sum contouring, using data from the Crystal City and Beeville Quadrangles. When a single variable such as uranium is of principal interest but is known to be related to other variables, weighted sums provide a simple way to incorporate the dependency.

One way to utilize both weighted sums and kriging would be to combine the separate contoured plots but the kriging variances generally would be large. The simpler technique of forming a new variable was used instead, termed Natural Factors. The variables incorporated and the weights are as follows

Permian

1.460* [L-U + (-2.040)] +

2.100* [L-SP + (-8.230)] +

0.610* [L-NA + (-4.540] +

0.610* [L-V + (4.540] +

0.920* [L-MO + (-1.910)]

Ogallala

1.58 • [L-U + (1.80)] +1.30 • [L-LI + (-4.52)] +

1.69 • [L-AJ + (-1.38)] +1.07 • [L-V + (-2.75)] +

1.51 • [L-MO + (-1.98)] + 2.09 • [L-MG + (-3.38)]

In each formation svg's were computed and plotted for the Natural Factors variable. These are shown in Figure 7. A functional model was fitted and used to compute the coefficients in the kriging estimator. Figure 8 shows the composite of the Ogallala/Permian Natural Factors kriged contours overlain with the Amaral favorable areas A, B, C, D and also those identified by Beauchamp and others (1980). Amaral area B is delineated clearly by one or more +4 contours. It is interesting to note that at approximately Lat. 34° 20' N and Long. 191° 10' W there are several +4 contours (Fig. 8). This corresponds to a pattern of large positive normalized residuals and high-level kriged uranium contours as shown in Figure 5. This region does not correspond directly to one of Amaral's areas. It does correspond to Area IIC. IIIC identified by Beauchamp and others (1980).

The Natural Factors contours do not seem to indicate any correspondence with Area D in the Permian. This is not unexpected because the favorable units in Area D are Pennsylvanian in age and are present only at greater depths and are not penetrated by the sampled wells.

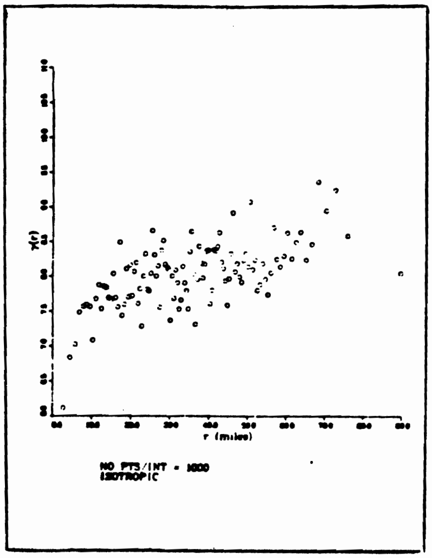
Another weighted sum, termed Subjective Mineralization also was kriged but was not as useful.

CONCLUSIONS

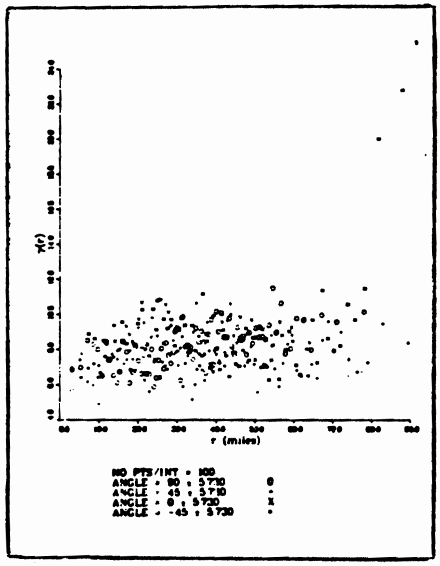
Kriging was determined to be a viable geostatistical tool for analyzing geochemical dispersion patterns. The variograms as estimated from the sample variogram plots clearly delineate between variables and geologic units and provide groupings naturally related to predicting Uranium occurrences.

As a tool to identify favorable areas for exploration for uranium several aspects of kriging were utilized; contour plots, normalized residuals, and weighted sum contouring. Coincidence was determined with two areas identified by Amaral (1979) and other areas identified by Beauchamp and others (1980).

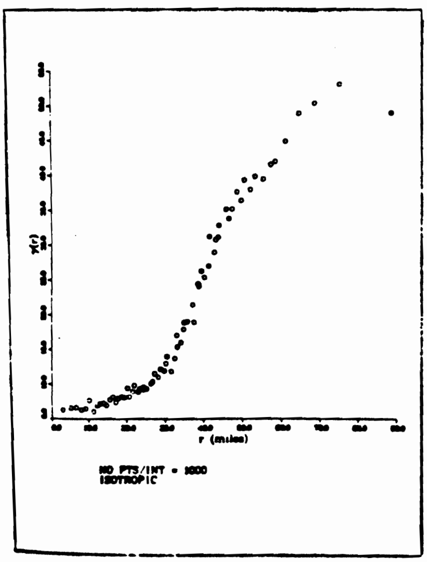
Kriging of linear combinations by forming a new variable, as was done with Natural Factors, is not optimal. Neither is kriging of each component. The optimal method is cokriging as is presented in Myers (1982, 1983, 1984) and Carr, Myers, and Glass (1985).



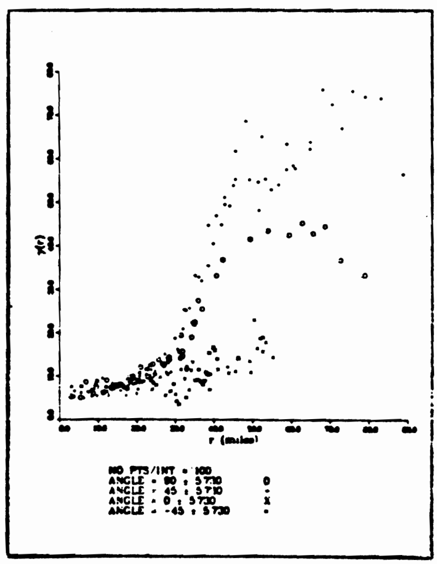
(a) Isotropic - Natural Factors



(b) Anisotropic - Natural Factors



(c) Isotropic - Natural Factors



(d) Anisotropic - Natural Factors

Figure 7. Weighted sum semivariograms for Permian units (a) and (b); weighted sum semivariograms for Ogallala Formation (c) and (d).

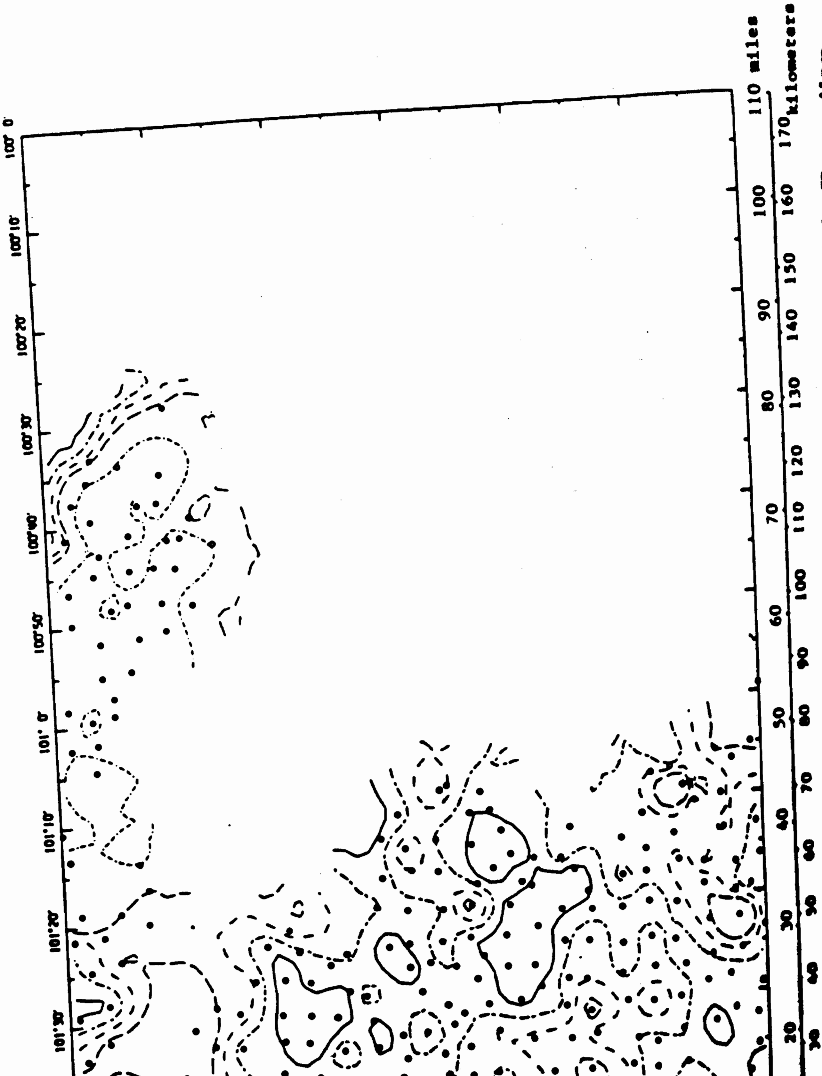


Figure 8-1. Weighted feature weighted mean contours from kriging for Ogallala Formation.

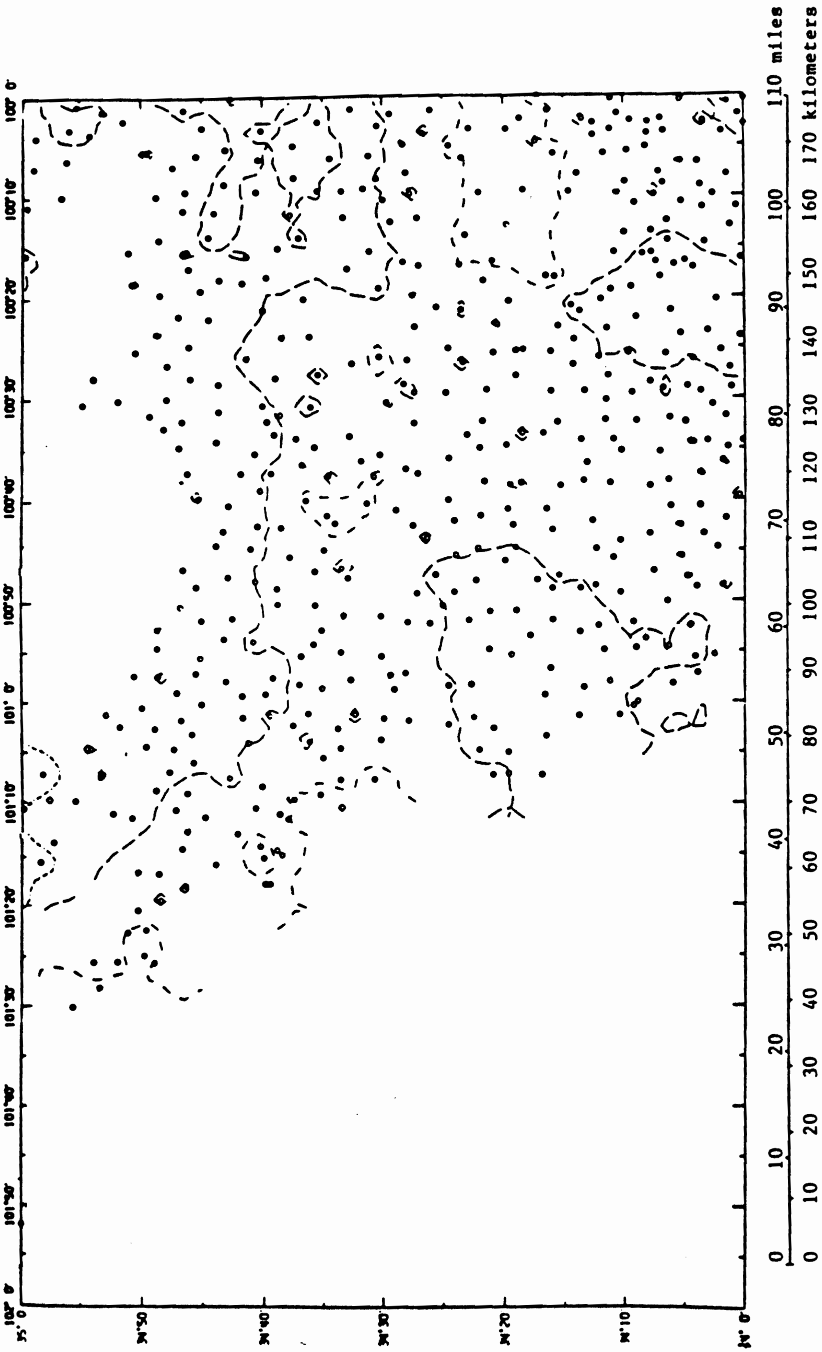
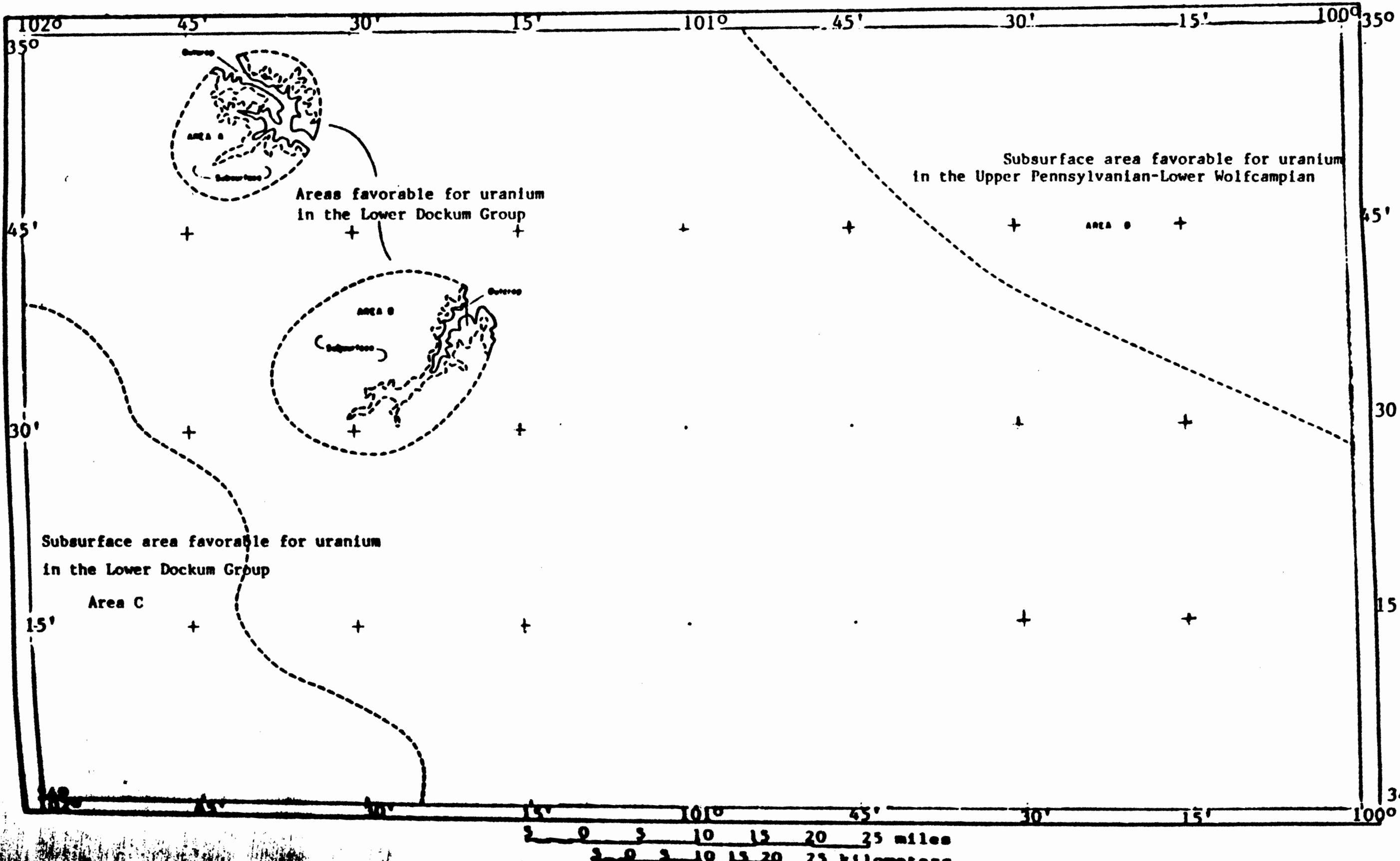


Figure 8-2. Natural factors weighted sum contours from kriging for Permian units.


102°
45'
30'
15'
101°
45'
30'
15'
100°
35°
34°
Outcrop
Area B
Subsurface
Areas favorable for uranium
in the Lower Dockum Group
Subsurface area favorable for uranium
in the Upper Pennsylvanian-Lower Wolfcampian
Subsurface area favorable for uranium
in the Lower Dockum Group
Area C
0 5 10 15 20 25 miles
0 5 10 15 20 25 kilometers

ACKNOWLEDGMENTS

This work was supported by the Oakridge Gaseous Diffusion Plant NURE project.

REFERENCES

Amaral, E. J., National Uranium Resource Evaluation Plainview Quadrangle: Bendix Field Engineering Corp., Grand Junction, Colorado, G. JQ-001(79), 26 p.

Beauchamp, J. J., Begovich, C. L., Kane, V. E., and Wolf, D. A., 1980, Application of discriminant analysis and generalized distance measures to uranium exploration: Jour. Math. Geology, v. 12, no. 6, p. 537-556.

Burgess, T. M., and Webster, R., 1980, Optimal interpolation and isarithmic mapping of soil properties I., The Semi-variogram and punctual kriging: Jour. Soil Science, v. 31, p. 315-331; II., Block kriging: Jour. Soil Science, v. 31, p. 533-541.

Carr J., Myers, D. E. , and Glass, C., 1985, Cokriging - a computer program: Computers & Geosciences, v. 11, no. 2, p. 111-127.

Everhart, D., 1977, Status and progress of the NURE program: Industry Seminar, U. S. Department of Energy, Grand Junction, Colorado, p. 69-102.

Journel, A. G., and Huijbrechts, Ch., 1978, Mining geostatistics: Academic Press, London, 600 p.

Kane, V. E. 1977, Geostatistics: Symposium on hydrogeochemical and stream sediment reconnaissance for uranium in the United States, March 16 and 17, 1977, U. S. Department of Energy, Grand Junction, Colorado. [GJBX-77(77)], p. 203-222.

Kane, V. E., Begovich, C. L. Butz, T. R. and Myers, D. E., 1982, Interpolation of regional geochemistry using optimal interpolation parameters: Computers & Geosciences, v. 8, no. 2, p. 117-136.

Matheron, G., 1965, Les variables regionalisees et leur estimation: Mason et Cie, Paris, 305 p.

Matheron, G., 1971, The theory of regionalized variables and its applications: Cahiers du Centre de Morphologie Mathematique de Fontainebleau, v. 5, 211 p.

Matheron, G., 1973, The intrinsic random functions and their applications: Advances Applied Probability, v. 5, p.437-468.

Myers, D. E., 1982, Matrix formulation of cokriging: Jour. Math. Geology, v. 14, no. 3, p. 249-257.

Myers, D. E., 1983 Estimation of linear combinations and co-kriging: Jour. Math. Geology, v. 13, no. 5, p. 633-637.

Myers, D. E., 1984, Cokriging-New developments, in Verly, G., and others, eds., Geostatistics for natural resource characterization: D. Reidel, Dordrecht, p. 295-305.

Myers, D. E., Begovich, C. L., Butz, T.R., and Kane , V. E., 1980, Application of kriging to hydrogeochemical data from the National Uranium Resource Evaluation Project: ORGDP, Oakridge, Tennessee, K/UR-44, 124 p.

Myers, D. E., Begovich, C. L., Butz, T. R., and Kane, V. E., 1982, Variogram models for regional geochemical data: Jour. Math. Geology, v. 14, no. 6, p. 629-644.

Roach, C., 1978, Possible NURE resource assessment methodologies: U. S. Department of Energy, Grand Junction, Colorado, 11 p.

Uranium Resource Evaluation Project, 1978, Hydrogeochemical and stream sediment reconnaissance basic data for Plainview NMTS Quadrangle, Texas: ORGDP, K/UR-101, 36 p.