UCRL- 84163
PREPRINT Rev. 1

MASTER

NUCLEAR WASTE REPOSITORY CHARACTERIZATION: A SPATIAL ESTIMATION/IDENTIFICATION APPROACH

J. V. Candy
N. Mao

This Paper Was Prepared For Submittal To
International Federation of Automatic
Control
VIII Triennial World Congress
Kyoto Japan, Aug. 24-28, 1981

March 11, 1981

NUCLEAR WASTE REPOSITORY CHARACTERIZATION: A SPATIAL ESTIMATION/IDENTIFICATION APPROACH*

J. V. Candy and N. Mao

Lawrence Livermore National Laboratory
Livermore, CA 94550

Abstract. In this paper we consider the application of spatial estimation techniques to a groundwater aquifer and geological borehole data. We investigate the adequacy of these techniques to reliably develop contour maps from various data sets.

The practice of spatial estimation is discussed and the estimator is then applied to a groundwater aquifer system and a deep geological formation. It is shown that the various statistical models must first be identified from the data and evaluated before reasonable results can be expected.

Keywords. Nuclear Waste Disposal, spatial estimation, geology, optimal filtering, identification, statistics.

INTRODUCTION

The need for alternate energy sources has propelled the use of nuclear energy as a feasible means for generating power. Nuclear technology has been utilized in many military and non-military applications such as fuel fabrication, medicine, weapons, etc. One perplexing problem has evolved - the disposal of radioactive waste products. It is the responsibility of the Nuclear Regulatory Commission (NRC) to assure that wastes be disposed in a manner non-detrimental to public health and welfare.

One proposed solution to the disposal problem is to isolate the waste in deep geological repositories. The waste is to be contained until they are no longer radiotoxic. Some elements will take a long time to decay - up to a few million years. For these slowly decaying elements the primary containment vessel will have deteriorated and therefore the geological structure of a repository will serve as the ultimate containment vessel. The question of contamination of the water supply becomes of prime concern. The magnitude of the disposal problem can be realized from the fact that only a single curie of ^{90}Sr, a common fission by-product, if dissolved in water

restrial storage have stressed the need for careful hydrogeologic studies to insure that the wastes will not contaminate existing or future water supplies. Thus, prior to construction and regulation of a repository, a site must be selected, characterized, and evaluated. Since the most likely release of radionuclides in the long term will be via the groundwater system, it is obvious that hydrogeology will form a major component of any performance evaluation model. Much information about the geological and hydrogeological structures of the site must be obtained. This process is called site characterization. This paper addresses itself to the site characterization problem, i.e., the problem of determining various geological and hydrogeological parameters necessary to characterize a potential repository. Information about hydrogeological parameters can be obtained by borehole sampling. However, too many boreholes at a site are not only expensive but they can tend to deteriorate the integrity of the site as a possible repository. Two approaches are possible: first, the use of nonintrusive measurements (e.g., seismic) or second, make statistical inferences on the suitability based on few measurements. In this paper we evaluate a technique which can be used to solve the site characterization

Abstract. In this paper we consider the application of spatial estimation techniques to a groundwater aquifer and geological borehole data. We investigate the adequacy of these techniques to reliably develop contour maps from various data sets.

The practice of spatial estimation is discussed and the estimator is then applied to a groundwater aquifer system and a deep geological formation. It is shown that the various statistical models must first be identified from the data and evaluated before reasonable results can be expected.

Keywords. Nuclear Waste Disposal, spatial estimation, geology, optimal filtering, identification, statistics.

INTRODUCTION

The need for alternate energy sources has propelled the use of nuclear energy as a feasible means for generating power. Nuclear technology has been utilized in many military and non-military applications such as fuel fabrication, medicine, weapons, etc. One perplexing problem has evolved - the disposal of radioactive waste products. It is the responsibility of the Nuclear Regulatory Commission (NRC) to assure that wastes be disposed in a manner non-detrimental to public health and welfare.

One proposed solution to the disposal problem is to isolate the waste in deep geological repositories. The waste is to be contained until they are no longer radiotoxic. Some elements will take a long time to decay - up to a few million years. For these slowly decaying elements the primary containment vessel will have deteriorated and therefore the geological structure of a repository will serve as the ultimate containment vessel. The question of contamination of the water supply becomes of prime concern. The magnitude of the disposal problem can be realized from the fact that only a single curie of ^{90}Sr, a common fission by-product, if dissolved in water could render 10^{11} liters unacceptable as drinking water according to the U.S. Public Health Service Standards (Davis, 1966). Disposal practices depend on the radioactivity and general chemical character of the waste, and the physical environment in the area of disposal. All plans for terrestrial storage have stressed the need for careful hydrogeologic studies to insure that the wastes will not contaminate existing or future water supplies. Thus, prior to construction and regulation of a repository, a site must be selected, characterized, and evaluated. Since the most likely release of radionuclides in the long term will be via the groundwater system, it is obvious that hydrogeology will form a major component of any performance evaluation model. Much information about the geological and hydrogeological structures of the site must be obtained. This process is called site characterization. This paper addresses itself to the site characterization problem, i.e., the problem of determining various geological and hydrogeological parameters necessary to characterize a potential repository. Information about hydrogeological parameters can be obtained by borehole sampling. However, too many boreholes at a site are not only expensive but they can tend to deteriorate the integrity of the site as a possible repository. Two approaches are possible: first, the use of nonintrusive measurements (e.g., seismic) or second, make statistical inferences on the suitability based on few measurements. In this paper we evaluate a technique which can be used to solve the site characterization problem. The fundamental problem is to generate a grid of control points from a set of sparse, irregular, uncertain, but spatially-correlated, measurements - this is called the fundamental spatial estimation problem.

*Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48.

he objective of the paper is to evaluate he effectiveness of spatial estimation as a otential tool for site characterization of eep geological repositories. The basic riterion for evaluation is how realistially the estimator predicts the geological henomenon. This can be judged in two ways: y comparing the estimated value with the easured value for the point which is reoved from the data set, or by comparing the stimated map with the "real" contour map ased on other information in addition to he estimated data set. We evaluate the stimator by experimenting on representative ata sets. The first case is a simple roundwater aquifer and the second is a omplex geological formation.

SPATIAL ESTIMATION

n this section we briefly state the fundaental theory of spatial estimation. Our ltimate goal is to use the spatial estiator to produce a reasonable contour map ith some measure of confidence. Here we evelop an "optimal" spatial estimator which an be used to estimate control points on a rid for eventual contouring.

et $z(\underline{x}_i)$ be the value of the variable z at oint $\underline{x}_i$ in space (1, 2, or 3 dimensional). n the estimation approach, $z(\underline{x}_i)$ is interpreted as a particular realization of a andom function $Z(\underline{x})$. More formally, $z(\underline{x}_i)$ s defined as a regionalized variable, i.e., function with spatial distribution which aries randomly from one location to anther. Spatial estimation theory is based n the observation that the variabilities of ll regionalized variables can be characterzed by a statistical model

e define the following measurement model or the data as

$$Z(\underline{x}) = r(\underline{x}) + m(\underline{x}) + \varepsilon(\underline{x}) \qquad (1)$$

here $r(\underline{x})$ is the residual value at location ; $m(\underline{x})$ is the systematic error or drift and $(\underline{x})$ is the random measurement error at $\underline{x}$. e also assume that the drift is a slowly ar,'ng function which can be approximated y a polynomial and the function Z is ocally stationary, i.e.,

$$\mathrm{Var}\{Z(\underline{x}+\underline{h}) - Z(\underline{x})\} =: 2\gamma(\underline{h}) \qquad (2)$$

here 2γ is the variogram of the difference.

GIVEN a set of measurement data $\{z(\underline{x}_i)\}$, $i = 1, \ldots, N$ and the measurement model

$$z(\underline{x}_k) = r(\underline{x}_k) + m(\underline{x}_k) + \varepsilon(\underline{x}_k), \qquad (3)$$

FIND the "best" linear, unbiased, minimum error variance estimate $\hat{z}(\underline{x}_k)$ of z at point $\underline{x}_k$.

The spatial estimation problem can be solved using optimization theory results. The result is (see Candy (1980) for details)

$$\hat{z}(\underline{x}_k) = \underline{\lambda}^T \underline{Z}, \text{ such that} \qquad (4)$$

$$\sigma_k^2 := \mathrm{Var}\{z(x_k) - \hat{z}(x_k)\} \text{ is minimum} \qquad (5)$$

The associated error variance is also estimated as

$$\sigma_k^2 = \underline{\theta}^T \underline{b}_k \qquad (6)$$

where θ is determined by the variogram and drift functions.

Thus, the technique generates point estimates $\hat{z}(\underline{x}_k)$ and a measure of the precision σ_k^2. In the next section we discuss the pragmatic issues in spatial estimation.

PRACTICE OF SPATIAL ESTIMATION

We discuss the problem of identifying a spatial model in this section. We describe properties and procedures to estimate the variogram from samples. Then we discuss some of the practical aspects of spatial estimation which can be used.

First we discuss the properties of the variogram. The variogram is more general than the usual covariance because in cases that the covariance does not exist the variogram does (e.g. Wiener process). The variogram is related to the covariance by

$$\gamma(\underline{h}) = C(0) - C(\underline{h})$$

where $C(\underline{h}) := \mathrm{Cov}(z(\underline{x}+\underline{h}), z(\underline{x}))$.

The variogram starts at 0^+, is a positive definite, even function, possesses a limit

SPATIAL ESTIMATION

n this section we briefly state the fundaental theory of spatial estimation. Our ltimate goal is to use the spatial estiator to produce a reasonable contour map ith some measure of confidence. Here we evelop an "optimal" spatial estimator which an be used to estimate control points on a rid for eventual contouring.

et $z(\underline{x}_i)$ be the value of the variable z at oint $\underline{x}_i$ in space (1, 2, or 3 dimensional). n the estimation approach, $z(\underline{x}_i)$ is inerpreted as a particular realization of a andom function $Z(\underline{x})$. More formally, $z(\underline{x}_i)$ s defined as a regionalized variable, i.e., function with spatial distribution which aries randomly from one location to anther. Spatial estimation theory is based n the observation that the variabilities of ll regionalized variables can be characterzed by a statistical model.

e define the following measurement model or the data as

$$Z(\underline{x}) = r(\underline{x}) + m(\underline{x}) + \varepsilon(\underline{x}) \qquad (1)$$

here $r(\underline{x})$ is the residual value at location ; $m(\underline{x})$ is the systematic error or drift and $(\underline{x})$ is the random measurement error at $\underline{x}$. e also assume that the drift is a slowly arying function which can be approximated y a polynomial and the function Z is ocally stationary, i.e.,

$$\mathrm{Var}\{Z(\underline{x}+\underline{h}) - Z(\underline{x})\} =: 2\gamma(\underline{h}) \qquad (2)$$

here 2γ is the variogram of the difference.

he residual is defined to be zero mean with tationary increments as in (2). The random easurement errors are assumed to be zero ean, uncorrelated with $z(\underline{x})$, and itself, ith covariance Σ.

sing the measurement model of (1) the patial estimation problem is

$$\sigma_k^2 := \mathrm{Var}\{z(x_k) - \hat{z}(x_k)\} \text{ is minimum} \qquad (5)$$

The associated error variance is also estimated as

$$\sigma_k^2 = \underline{\theta}^T \underline{b}_k \qquad (6)$$

where θ is determined by the variogram and drift functions.

Thus, the technique generates point estimates $\hat{z}(\underline{x}_k)$ and a measure of the precision σ_k^2. In the next section we discuss the pragmatic issues in spatial estimation.

PRACTICE OF SPATIAL ESTIMATION

We discuss the problem of identifying a spatial model in this section. We describe properties and procedures to estimate the variogram from samples. Then we discuss some of the practical aspects of spatial estimation which can be used.

First we discuss the properties of the variogram. The variogram is more general than the usual covariance because in cases that the covariance does not exist the variogram does (e.g. Wiener process). The variogram is related to the covariance by

$$\gamma(\underline{h}) = C(0) - C(\underline{h})$$

where $C(\underline{h}) := \mathrm{Cov}(z(\underline{x}+\underline{h}), z(\underline{x}))$.

The variogram starts at $0^{\dagger}$, is a positive definite, even function, possesses a limit

---

† In some cases, the variogram may not start at zero, but some other positive value then it is said to be discontinuous and exhibit the "nugget effect" (see (Journel, 1978) for details), i.e., $\gamma(\underline{h}) = C_0$ for $\underline{h} > \underline{\varepsilon}$ and $\underline{\varepsilon}$ small.

called the sill[†] (C(0)), a range, the point the say h = a where the samples are uncorrelated, and the variogram grows more slowly than a parabola in h^2 (see Journel, 1978 for details).

Table I depicts some of the common variograms which satisfy the properties discussed. Note that some variograms do not possess all of these properties (e.g. a linear variogram does not possess a sill). Finally we note that if the regionalized variables do not exhibit the same behavior in every direction (isotropy), they are called anisotropic, i.e., the variograms calculated in different spatial directions differ. When they are identical, the variogram is termed isotropic.

Variogram identification from raw data is the most crucial part of the spatial estimation process. Prior to the actual estimation, it is necessary to fit an experimental variogram to a theoretical model which will ensure mathematical consistency of the calculations. Thus, the practice of variogram identificaion is concerned with:

(i) estimating the "raw" or experimental variogram and drift from sample data;
(ii) fitting a theoretical variogram and drift to the experimental; and
(iii) checking the validity of the fit.

An unbiased estimator for the variogram is (Olea, 1975)

$$\hat{\gamma}^*(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{i=1}^{N(\underline{h})} (z(\underline{x}_i+\underline{h}) - z(\underline{x}_i))^2 \quad (7)$$

where $N(\underline{h})$ is the number of pairs of points separated by distance $\underline{h}$.

If drift is present, then it can be shown that (15) is a biased estimator for the variogram. The drift can also be estimated using the experimental drift estimator on the sample data, i.e.,

$$\hat{m}^*(\underline{h}) = \frac{1}{N(\underline{h})} \sum_{i=1}^{N(\underline{h})} (z(\underline{x}_i+\underline{h}) - z(\underline{x}_i)) \quad (8)$$

The experimental variogram can then be drift-corrected and fit. The drift corrected variogram is given by

$$\hat{\gamma}(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{i=1}^{N(\underline{h})} [z(\underline{x}_i+\underline{h}) - z(\underline{x}_i) - \hat{m}^*(\underline{h})]^2$$

algebraic property that higher order differencing of variables filters out polynomials in the expectation. The original random function $Z(\underline{x})$ is called an intrinsic random function of order k (abbreviated k-IRF) where k is the highest degree polynomial filtered. The purpose of taking increments is to produce a stationary regionalized variable from one with a drift. The advantage is that the covariance structure of the spatial variable can be estimated without the effects of the drift. This is done by developing the generalized covariance of the k-IRF which differs from the variogram, which is only legitimate for the 0-IRF case.

There are various classes of functions that satisfy the conditions of a generalized covariance, but one class with nice properties for identification purposes (that is, linear in the coefficients) is the class of polynomial generalized covariances. The form of these generalized covariances, which depends on the order of the increment, is listed in Table I.

Once we have decided on the type of a theoretical variogram or generalized covariance, we must make sure that the fit is reasonable. Tests must be performed to insure the validity of the variogram and drift models (Gambolati, 1979). The technique employed is the successive estimation of all of the data points, ignoring each of them in turn, one by one. We then verify that there is no systematic error and compare the calculated errors (difference between the estimated and real values) with the theoretically predicted standard deviations (σ_k) to assure consistency. A statistical analysis is carried out on the standard errors, checking that they are zero mean and unit variance. This technique can be utilized for comparing several models and determining the best fit. The first selection criterion is minimum mean squared error and secondly the standard error is close to unity. The deviation from the actual values, i.e., $\Delta z_i = z_i - \hat{z}_i$ are calculated for the entire data set and then the sample error statistics are estimated. The sample statistics are:

(i) Systematic Error is zero. ($E(\Delta z_i) = 0$);
(ii) Standard Error is unit variance. ($\frac{\Delta z_i}{\sigma_i} \sim N(0,1)$);

linear variogram does not possess a sill). Finally we note that if the regionalized variables do not exhibit the same behavior in every direction (isotropy), they are called anisotropic, i.e., the variograms calculated in different spatial directions differ. When they are identical, the variogram is termed isotropic.

Variogram identification from raw data is the most crucial part of the spatial estimation process. Prior to the actual estimation, it is necessary to fit an experimental variogram to a theoretical model which will ensure mathematical consistency of the calculations. Thus, the practice of variogram identificaion is concerned with:

(i) estimating the "raw" or experimental variogram and drift from sample data;
(ii) fitting a theoretical variogram and drift to the experimental; and
(iii) checking the validity of the fit.

An unbiased estimator for the variogram is (Olea, 1975)

$$\hat{\gamma}^*(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{i=1}^{N(\underline{h})} (z(\underline{x}_i+\underline{h}) - z(\underline{x}_i))^2 \quad (7)$$

where $N(\underline{h})$ is the number of pairs of points separated by distance $\underline{h}$.

If drift is present, then it can be shown that (15) is a biased estimator for the variogram. The drift can also be estimated using the experimental drift estimator on the sample data, i.e.,

$$\hat{m}^*(\underline{h}) = \frac{1}{N(\underline{h})} \sum_{i=1}^{N(\underline{h})} (z(\underline{x}_i+\underline{h}) - z(\underline{x}_i)) \quad (8)$$

The experimental variogram can then be drift-corrected and fit. The drift corrected variogram is given by

$$\hat{\gamma}(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{i=1}^{N(\underline{h})} [z(\underline{x}_i+\underline{h}) - z(\underline{x}_i) - \hat{m}^*(\underline{h})]^2$$

Another approach to the drift and structure identification problem is to make use of an out the effects of the drift. This is done by developing the generalized covariance of the k-IRF which differs from the variogram, which is only legitimate for the 0-IRF case.

There are various classes of functions that satisfy the conditions of a generalized covariance, but one class with nice properties for identification purposes (that is, linear in the coefficients) is the class of polynomial generalized covariances. The form of these generalized covariances, which depends on the order of the increment, is listed in Table I.

Once we have decided on the type of a theoretical variogram or generalized covariance, we must make sure that the fit is reasonable. Tests must be performed to insure the validity of the variogram and drift models (Gambolati, 1979). The technique employed is the successive estimation of all of the data points, ignoring each of them in turn, one by one. We then verify that there is no systematic error and compare the calculated errors (difference between the estimated and real values) with the theoretically predicted standard deviations (σ_k) to assure consistency. A statistical analysis is carried out on the standard errors, checking that they are zero mean and unit variance. This technique can be utilized for comparing several models and determining the best fit. The first selection criterion is minimum mean squared error and secondly the standard error is close to unity. The deviation from the actual values, i.e., $\Delta z_i = z_i - \hat{z}_i$ are calculated for the entire data set and then the sample error statistics are estimated. The sample statistics are:

(i) Systematic Error is zero. $(E(\Delta z_i) = 0)$;
(ii) Standard Error is unit variance. $\left(\frac{\Delta z_i}{\sigma_i} \sim N(0,1)\right)$;
(iii) RMS Error is minimum $(\min \sqrt{\Delta z_i^2})$

This completes the discussion of variograms estimation and identification.

APPLICATION OF SPATIAL ESTIMATORS

In this section we discuss the application of the spatial estimation techniques to real

† When the variance exists.

data sets. We apply the "practice" of spatial estimation to representative measurements of water and geological formation depth data in order to evaluate the performance of the estimator and produce contour maps for comparative purposes. We evaluate the estimator performance on a groundwater aquifer system and then investigate the subsurface structure mapping of a geological formation from borehole data. The aquifer system structure is simple, while the rock depth information is much more complex. First we investigate an aquifer located in Toppenish Creek, Washington.

CASE STUDY: Toppenish Creek Basin, Washington

The purpose of this study is to investigate the performance of spatial estimation techniques on a simple groundwater aquifer system. The data set is 76 measurements of the September 1971 water levels at Toppenish Creek Basin, Washington. The measurements are distributed irregularly in the region and contours of the raw data show that the major hydraulic gradient is from the SE to NW direction. The spatial estimator will be applied to generate a grid of (22 x 22) uniformly spaced sample estimates and then contoured.

First, we must identify a structural model (variogram and drift) from the data. Following the procedure outlined in the previous section the estimated variograms $\hat{\gamma}^*$, $\hat{\gamma}$ and drift $\hat{m}^*$ are calculated. A set of sample anisotropic/isotropic variograms are estimated and as expected the NW/SE variogram (direction of the hydraulic gradient) increased at a rate faster than h^2 indicating the presence of a drift. Note that the average variogram is dominated by the NW/SE and N/S variogram. An examination of the estimated drift-corrected variograms (Fig. 1) still indicate a biased estimate (not all drift removed); therefore we decided to visually fit the NE/SW variogram, since it should be essentially drift free (appears perpendicular to the major hydraulic gradient).

The estimated NE/SW variogram is shown in Fig. 2. We noted that a large number of pairs (~ 120) are at distances between 0 and 2 feet, which indicates that a careful fit of the initial points should be made, since function, linear, and spherical variograms with constant and linear drifts. After fitting these structures visually we applied the successive estimation method for validation. Each model was fit visually to the estimated NE/SW variogram and then the parameters of each particular structural model were adjusted until the standard error was approximately unity. Those with the smallest root mean-squared error (RMSE) were retained. These results are summarized in Tables II and III. We found that the successive estimation technique appears to identify the closest points with most pairs. This is apparent when we examine Fig. 2 more closely which shows the "best" fits from Tables II and III: generalized covariance, spherical, and linear variograms with constant drift. Note that in this figure we plot the "fits" to the NE/SW variogram since it is approximately drift-free. Note also that the initial slopes of the linear and spherical are practically identical, thus they yield similar validations. We also found that most pairs (range, sill) with initial slope 25 yielded identical statistics again confirming the heavy dependence on closely spaced pairs.

After the structural identification was completed, the spatial estimator was run over the data set. We chose the linear variogram with a constant drift since it had the minimum RMS error and generated the contours shown in Figs. 3 and 4. We could have selected the generalized covariance, or spherical model as well, since the validation statistics were very close. In fact, maps generated from these models were almost identical. An examination of the resulting contour map of the Toppenish Creek Basin indicates that most of the major features have been maintained. Note that the normalized map coordinates must be converted back to problem coordinates and the conversion factors are shown as (Δx, Δy) on Figs. 3 and 4. A close examination of the one-sigma error map in Fig. 4 indicates that the upper NE corner and lower SW corners are the most uncertain areas. This is expected since no measurements are available in those regions.

We also note that the spatial estimator was evaluated for another aquifer (Todd Lake, Penn.) in which the true grid values of the region were assumed known. Data (28 points) was selected† from an existing "truth" model (partial differential equation) of the

Case Study: Toppenish Creek Basin, Washington

The purpose of this study is to investigate the performance of spatial estimation techniques on a simple groundwater aquifer system. The data set is 76 measurements of the September 1971 water levels at Toppenish Creek Basin, Washington. The measurements are distributed irregularly in the region and contours of the raw data show that the major hydraulic gradient is from the SE to NW direction. The spatial estimator will be applied to generate a grid of (22 x 22) uniformly spaced sample estimates and then contoured.

First, we must identify a structural model (variogram and drift) from the data. Following the procedure outlined in the previous section the estimated variograms $\hat{\gamma}^*$, $\hat{\gamma}$ and drift $\hat{m}^*$ are calculated. A set of sample anisotropic/isotropic variograms are estimated and as expected the NW/SE variogram (direction of the hydraulic gradient) increased at a rate faster than h^2 indicating the presence of a drift. Note that the average variogram is dominated by the NW/SE and N/S variogram. An examination of the estimated drift-corrected variograms (Fig. 1) still indicate a biased estimate (not all drift removed); therefore we decided to visually fit the NE/SW variogram, since it should be essentially drift free (appears perpendicular to the major hydraulic gradient).

The estimated NE/SW variogram is shown in Fig. 2. We noted that a large number of pairs (~ 120) are at distances between 0 and 2 feet, which indicates that a careful fit of the initial points should be made, since the estimator will weight these closest points most heavily. The drift appears linear; however, it could be modeled either as linear or a constant since it is not very severe. We used the automatic (generalized covariance) fit feature of the estimation algorithm† and also selected power closely which shows the best fits from Tables II and III: generalized covariance, spherical, and linear variograms with constant drift. Note that in this figure we plot the "fits" to the NE/SW variogram since it is approximately drift-free. Note also that the initial slopes of the linear and spherical are practically identical, thus they yield similar validations. We also found that most pairs (range, sill) with initial slope 25 yielded identical statistics again confirming the heavy dependence on closely spaced pairs.

After the structural identification was completed, the spatial estimator was run over the data set. We chose the linear variogram with a constant drift since it had the minimum RMS error and generated the contours shown in Figs. 3 and 4. We could have selected the generalized covariance, or spherical model as well, since the validation statistics were very close. In fact, maps generated from these models were almost identical. An examination of the resulting contour map of the Toppenish Creek Basin indicates that most of the major features have been maintained. Note that the normalized map coordinates must be converted back to problem coordinates and the conversion factors are shown as (Δx, Δy) on Figs. 3 and 4. A close examination of the one-sigma error map in Fig. 4 indicates that the upper NE corner and lower SW corners are the most uncertain areas. This is expected since no measurements are available in those regions.

We also note that the spatial estimator was evaluated for another aquifer (Todd Lake, Penn.) in which the true grid values of the region were assumed known. Data (28 points) was selected† from an existing "truth" model (partial differential equation) of the aquifer. The variogram and drift models were identified from the sparse data. The objective was to use these models based on sparse data to

† We used the algorithm BLUEPACK (Delfiner, 1976) for our case studies.

† An LLNL hydrologist, who developed the truth model, selected the data points. Ten were his original measurements and the other 18 he obtained from knowledge (maps, surveys, etc.) of the region.

estimate the "truth" grid. Comparison of the grids showed that the estimator could reproduce the truth grid (323 points) with a maximum relative error of 1% (see Candy (1980) for details).

Summarizing the results of these case studies, it appears that the spatial estimator can be used with confidence to reconstruct a regional contour map of simple groundwater aquifer systems. In the next section we consider the mapping of a deep geological structure.

CASE STUDY: Nevada Test Site, Paleozoic Rock Formations

The purpose of this study is to evaluate the performance of the spatial estimator on data from a complex but well known geological formation - the Paleozoic rock at the Nevada Test Site (NTS). The mile NTS is one of the most exhaustively studied areas in existence. We studied the depth to the top of the Paleozoic rocks from 93 borehole data and the effect of a fault on the estimated depth. First, the variogram was estimated from the raw data. A set of estimated anisotropic (directional) variograms are shown in Fig. 5. These variograms are biased because of the presence of drift in all directions.

Various tested structural models are summarized in Table IV and the identification and validation results are summarized in Table V. The RMS errors for those models without fault are slightly smaller than those for the equivalent models with fault. There are probably two reasons behind this observation. First, the vertical throw due to fault is relatively small as compared with the error. Second, there are fewer data points to estimate because the fault effectively divides the data into two sets and no correlation is assumed between them. Nevertheless, the RMS error from all models are not too different from the RMS error of 77.8 m estimated from gravity data based on 20 measurements of depth between 500 and 1000 m which validates the results.

The contour maps based on the estimated depth for the case with the fault together with the corresponding uncertainty maps are shown in Figs. 6 and 7. Although the estimated depths show discontinuity near the fault, the contour routine smooths out these residuals (see Table V) using the same initial variogram but constant drift is quite good (0.941) indicating that the estimated drift is reasonable.

The structure for the rock depth is much more complicated than those from the water level case studied before. However, with simple linear variogram and constant drift, the Paleozoic rock depth can be reasonably reproduced.

This concludes the application of spatial estimation techniques to a simple groundwater aquifer system and complex geological formations.

SUMMARY

In this paper we developed spatial estimators for correlated data irregularly distributed in a region. The practice of spatial estimation was then discussed and it was shown that the most crucial area of spatial estimation is the identification of the statistical structural model - the variogram and drift. Techniques to estimate and identify the underlying variogram and drift from raw data were discussed and a validation technique discussed.

The spatial estimators were applied to two case studies: (i) Toppenish Creek Basin, and (ii) Nevada Test Site. The first case was a simple groundwater aquifer and the second a complex geological formation. The spatial estimator yielded reasonable contour maps of the regions under investigation. Results of the estimator for the Nevada Test Site data were confirmed by using additional information (gravity measurements). Spatial estimation techniques can be an effective tool to characterize proposed repository sites from limited data (see Candy (1980)); however, much care must be taken when identifying the structural model from the raw data.

ACKNOWLEDGEMENTS

We would like to thank Mr. D. Freeman for helping us with much of the computer work done in this report. We would also like to thank Mr. J. Skrivan of USGS for his algorithm, Toppenish Creek data set and many technical discussions on this subject. We would finally but by no means least like to thank Ms. L. Lopez for typing the

The purpose of this study is to evaluate the performance of the spatial estimator on data from a complex but well known geological formation - the Paleozoic rock at the Nevada Test Site (NTS). The mile NTS is one of the most exhaustively studied areas in existence. We studied the depth to the top of the Paleozoic rocks from 93 borehole data and the effect of a fault on the estimated depth. First, the variogram was estimated from the raw data. A set of estimated anisotropic (directional) variograms are shown in Fig. 5. These variograms are biased because of the presence of drift in all directions.

Various tested structural models are summarized in Table IV and the identification and validation results are summarized in Table V. The RMS errors for those models without fault are slightly smaller than those for the equivalent models with fault. There are probably two reasons behind this observation. First, the vertical throw due to fault is relatively small as compared with the error. Second, there are fewer data points to estimate because the fault effectively divides the data into two sets and no correlation is assumed between them. Nevertheless, the RMS error from all models are not too different from the RMS error of 77.8 m estimated from gravity data based on 20 measurements of depth between 500 and 1000 m which validates the results.

The contour maps based on the estimated depth for the case with the fault together with the corresponding uncertainty maps are shown in Figs. 6 and 7. Although the estimated depths show discontinuity near the fault, the contour routine smooths out these differences. The contour map has the gross features of a map based on gravity data, surface geology data as well as the borehole data.

The contour maps indicate a quadratic drift. We calculated this drift with a spherical variogram of range 3,000 m and sill 16,800 m. From the variograms of the drift removed residuals it was evident that the anisotropy has been greatly reduced. The standard error for the drift removed

SUMMARY

In this paper we developed spatial estimators for correlated data irregularily distributed in a region. The practice of spatial estimation was then discussed and it was shown that the most crucial area of spatial estimation is the identification of the statistical structural model - the variogram and drift. Techniques to estimate and identify the underlying variogram and drift from raw data were discussed and a validation technique discussed.

The spatial estimators were applied to two case studies: (i) Toppenish Creek Basin, and (ii) Nevada Test Site. The first case was a simple groundwater aquifer and the second a complex geological formation. The spatial estimator yielded reasonable contour maps of the regions under investigation. Results of the estimator for the Nevada Test Site data were confirmed by using additional information (gravity measurements). Spatial estimation techniques can be an effective tool to characterize proposed repository sites from limited data (see Candy (1980)); however, much care must be taken when identifying the structural model from the raw data.

ACKNOWLEDGEMENTS

We would like to thank Mr. D. Freeman for helping us with much of the computer work done in this report. We would also like to thank Mr. J. Skrivan of USGS for his algorithm, Toppenish Creek data set and many technical discussions on this subject. We would finally but by no means least like to thank Ms. L. Lopez for typing the manuscript.

LIST OF REFERENCES

Candy, J. V. and N. Mao (1980) "Site Characterization: A Spatial Estimation Approach," Lawrence Livermore National Laboratory Report, UCID-18688.

Davis, S. N. and R. J. DeWiest (1966) Hydrogeology, Wiley, New York,.

Delfiner, P., J. P. Delhome, and J. P. Chiles (1976) BLUEPACK, Ecole Nationale Superieure Des Mines de Paris.

Gambolati, G. and G. Volpi (1979) "Groundwater Contour Mapping in Venice by Stochastic Interpolators 1. Theory," Water Resources Research, Vol. 15.

Journel, A. G. and Ch. J. Huijbregts (1978) Mining Geostatistics, Academic Press.

Matheron, G. (1971) "The Theory of Regionalized Variables and Its Applications," Les Cahiers du Centre de Morphologie Mathematique, n°5, Ecole Nat. Sup. des Mines de Paris.

Olea, R. A. (1975) "Optimum Mapping Techniques Using Regionalized Variable Theory," Kansas Geological Survey.

Skrivan, J. A. and M. R. Karlinger (1980) "Semi-Variogram Estimation and Universal Kriging Program," Water Resources Division, USGS Report, Tacoma, Washington.

TABLE I Variogram and generalized covariance models

TYPE	MODEL
Variograms:	
Nugget effect	$\gamma(h) = C(1-\delta)$
Power	$\gamma(h) = a\lvert h\rvert^{\lambda}$
Spherical	$\gamma(h) = \begin{cases} \frac{C}{2}\left[\frac{3}{2}\frac{\lvert h\rvert}{a} - \frac{1}{2}\left(\frac{\lvert h\rvert}{a}\right)^3\right] & h \le a \\ C & h > a \end{cases}$
Gaussian	$\gamma(h) = a\left(1-\exp(-\lvert h\rvert^2/a^2)\right)$
Exponential	$\gamma(h) = a\left(1-\exp(-\lvert h\rvert/a)\right)$
Hyperbolic or cardinal sine	$\gamma(h) = 2a\,\ln(\lvert h\rvert)$
Generalized Covariances	
DRIFT — Constant (0)	POLYNOMIAL GENERALIZED COVARIANCE MODEL* $K(\lvert h\rvert) = C_0 + b_0\lvert h\rvert$
Linear (1)	$K(\lvert h\rvert) = C_0 + b_0\lvert h\rvert + b_1\lvert h\rvert^3$
Quadratic (2)	$K(\lvert h\rvert) = C_0 + b_0\lvert h\rvert + b_1\lvert h\rvert^3 + b_2\lvert h\rvert^5$

*Beta coefficient constraints:

(Two-dimensional space):

$C \ge 0,\ b_0 \ge 0,\ b_2 \ge 0,\ b_1 \ge -(10/3)\sqrt{b_0 b_2}$;

(Three-dimensional space):

$C \ge 0,\ b_0 \ge 0,\ b_2 \ge 0,\ b_1 \ge -\sqrt{b_0 b_2}$.

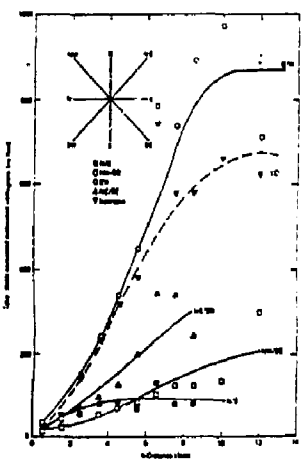


Fig. 1. Drift-corrected anisotropic/isotropic estimated variograms for Toppenish Creek data.

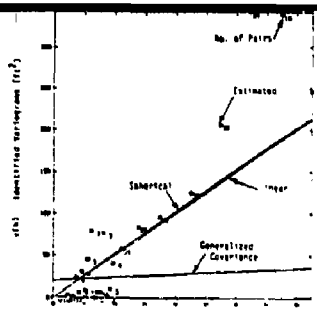


Fig. 2. Identified variograms (near origin) against NE/SW estimated variograms for Toppenish Creek Basin, Washington

TABLE II Toppenish Creek structural model summary

TYPE	MODEL	DRIFT
GENERALIZED COVARIANCE (AUTO)	$K(h) = 20.54 - 1.76\lvert h\rvert$	Linear
VARIOGRAM 1	$\gamma(h) = 95\lvert h\rvert^{1.7}$	Constant
2	Same	Linear
VARIOGRAM 3	$\gamma(h) = 600\left(3/2\,\frac{\lvert h\rvert}{35} - 1/2\left(\frac{\lvert h\rvert}{35}\right)^3\right)$	Constant
4	Same	Linear
VARIOGRAM 5	$\gamma(h) = 25\lvert h\rvert$	Constant

TABLE III Toppenish Creek structure identification/validation

TYPE	STANDARD ERROR	ROOT MEAN SQUARED ERROR (ft)	MEAN ERROR (ft)
GENERALIZED COVARIANCE (AUTO)	0.84	4.77	0.03
VARIOGRAM 1	0.96	5.18	0.15
2	0.96	5.39	0.16
VARIOGRAM 3	0.99	4.67	0.05
4	0.98	4.97	0.09
VARIOGRAM 5	1.02	4.66	0.06

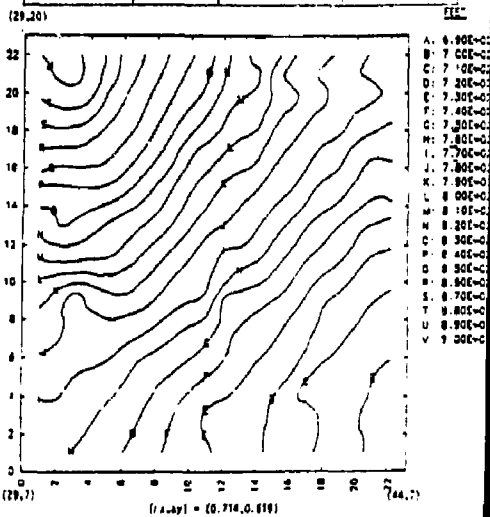


Fig. 3. Contour of groundwater levels (in feet) at Toppenish Creek Basin, Washington, using spatial estimation techniques (linear variogram)

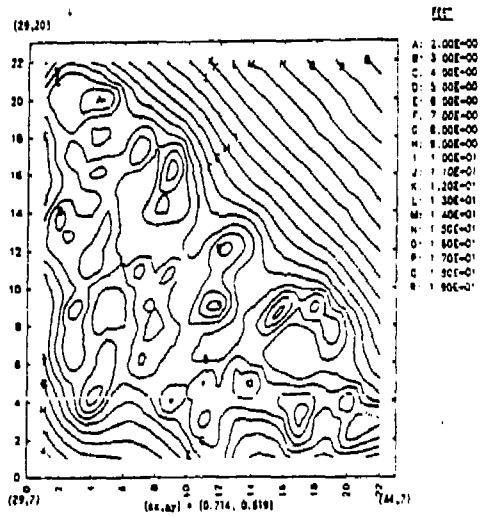


Fig. 4. One sigma estimation error contours (in feet) for Toppenish Creek Basin, Washington using linear variogram

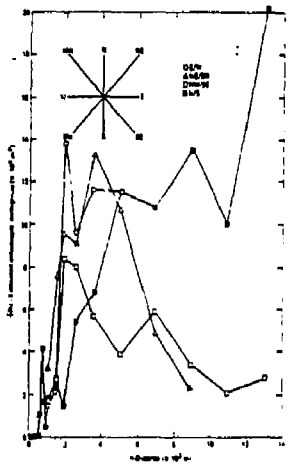


Fig. 5. Anisotropic variograms for paleozoic rock depth at NTS.

TABLE IV NTS paleozoic rock depth structural model summary

TYPE	MODEL	DRIFT
WITH FAULT		
Generalized Covariance (Auto)	$\gamma(h) = -4.4115\|h\| + 2.1088 \times 10^{-4}\|h\|^3$	Linear
Linear 1	$\gamma(h) = 9.15\|h\|$	Constant
Linear 2	$\gamma(h) = 9.15\|h\|$	Linear
Spherical A1	$\gamma(h) = 11700\left[\frac{3}{2}\frac{\|h\|}{2000} - \frac{1}{2}\left(\frac{\|h\|}{2000}\right)^3\right]$	Constant
Spherical A2	$\gamma(h) = 11900\left[\frac{3}{2}\frac{\|h\|}{2000} - \frac{1}{2}\left(\frac{\|h\|}{2000}\right)^3\right]$	Linear
WITHOUT FAULT		
Generalized		

TABLE V NTS paleozoic rock structure identification/validation

TYPE	STANDARDIZED ERROR	RMS ERROR (m)	MEAN ERROR (m)
WITH FAULT			
Generalized Covariance (Auto)	1.756	84	1.11.
Linear 1	0.999	76	1.288
Linear 2	1.023	79	1.733
Spherical A1	1.009	76	0.386
Spherical A2	1.009	80	-0.374
WITHOUT FAULT			
Generalized Covariance (Auto)	0.991	83	-5.234
Linear 1	1.003	73	-2.089
Linear 2	1.012	74	-5.464
Linear 3	1.010	77	0.505
Spherical A1	1.004	75	0.813
Spherical A2	1.003	73	-6.7..
Spherical A3	1.005	78	-0.373
Spherical d3	1.075	79	0.36.
Spherical C1=	0.941	69	-0.4..

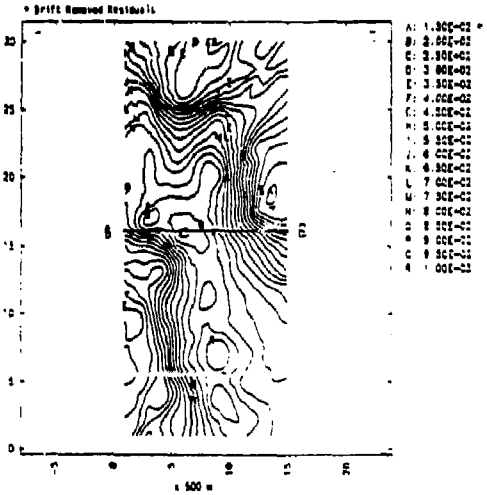
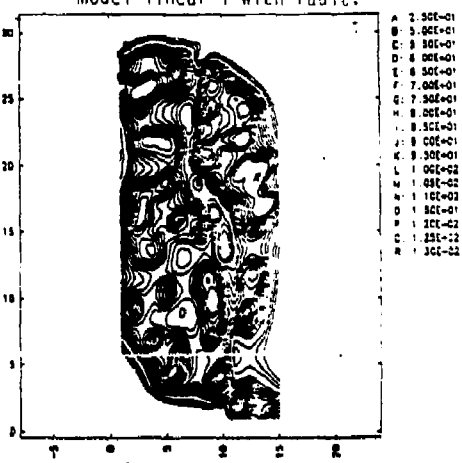


Fig. 6. Paleozoic rock depth at NTS model linear 1 with fault.



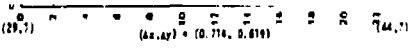


Fig. 4. One sigma estimation error contours (in feet) for Toppenish Creek Basin, Washington using linear variogram

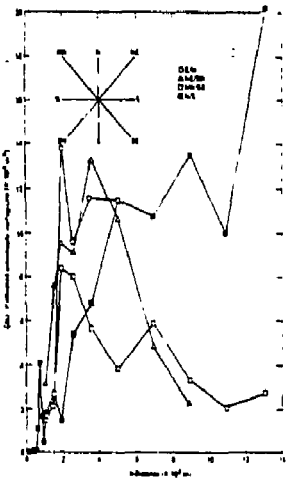


Fig. 5. Anisotropic variograms for paleozoic rock depth at NTS.

TABLE IV NTS paleozoic rock depth structural model summary

TYPE	MODEL	DRIFT
WITH FAULT		
Generalized Covariance (Auto)	$\gamma(h) = -4.8115\|h\| + 2.1088 \times 10^{-6}\|h\|^3$	Linear
Linear 1	$\gamma(h) = 9.15\|h\|$	Constant
Linear 2	$\gamma(h) = 9.15\|h\|$	Linear
Spherical A1	$\gamma(h) = 11700\ \frac{3}{2}\frac{\|h\|}{2000} - \frac{1}{2}\left(\frac{\|h\|}{2000}\right)^3$	Constant
Spherical A2	$\gamma(h) = 11900\ \frac{3}{2}\frac{\|h\|}{2000} - \frac{1}{2}\left(\frac{\|h\|}{2000}\right)^3$	Linear
WITHOUT FAULT		
Generalized Covariance (Auto)	$\gamma(h) = -5.0865\|h\| + 2.7081 \times 10^{-6}\|h\|^3$	Linear
Linear 1	$\gamma(h) = 8.70\|h\|$	Constant
Linear 2	$\gamma(h) = 8.70\|h\|$	Linear
Linear 3	$\gamma(h) = 9.15\|h\|$	Quadratic
Spherical A1	$\gamma(h) = 11100\ \frac{3}{2}\frac{\|h\|}{2000} - \frac{1}{2}\left(\frac{\|h\|}{2000}\right)^3$	Constant
Spherical A2	$\gamma(h) = 10800\ \frac{3}{2}\frac{\|h\|}{2000} - \frac{1}{2}\left(\frac{\|h\|}{2000}\right)^3$	Linear
Spherical A3	$\gamma(h) = 11250\ \frac{3}{2}\frac{\|h\|}{2000} - \frac{1}{2}\left(\frac{\|h\|}{2000}\right)^3$	Quadratic
Spherical B3	$\gamma(h) = 16800\ \frac{3}{2}\frac{\|h\|}{3000} - \frac{1}{2}\left(\frac{\|h\|}{3000}\right)^3$	Quadratic
Spherical C1*	$\gamma(h) = 16800\ \frac{3}{2}\frac{\|h\|}{3000} - \frac{1}{2}\left(\frac{\|h\|}{3000}\right)^3$	Constant

*Drift Removed Residuals

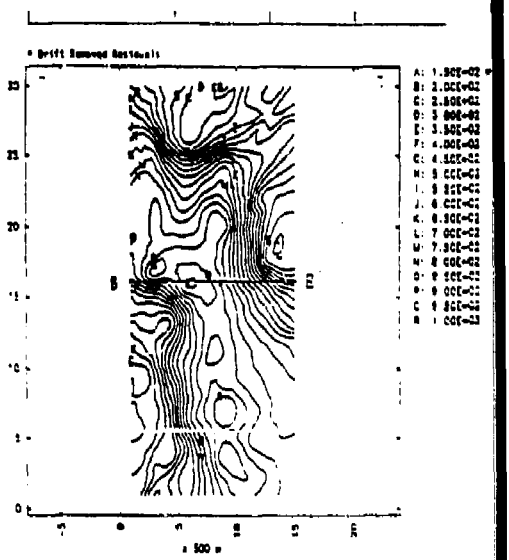


Fig. 6. Paleozoic rock depth at NTS model linear 1 with fault.

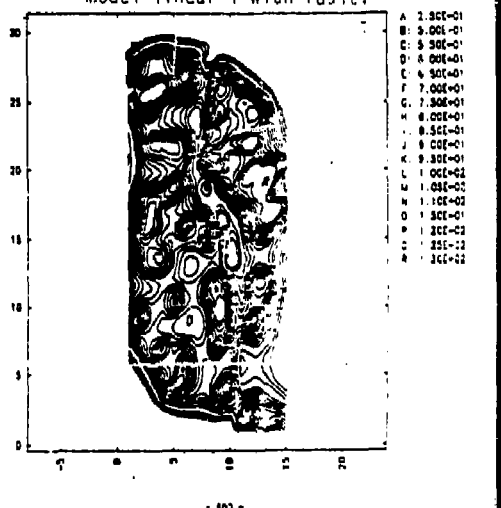


Fig. 7. One-sigma error for paleozoic rock depth at NTS from model linear 1 with fault.