with W.V. Harper (senior author), "Assessment of uncertainty in a travel path -- a case study in multivariate simulation" in P.A. Dowd et J.J. Royer (eds), 2nd CODATA Conference on Geomathematics and Geostatistics. *Sci. de la Terre, Sér. Inf.*, Nancy 1992, vol 31, pp.317 − 327

Assessment of uncertainty in a travel path - a case study in multivariate simulation

William V. Harper
Resource International
281 Enterprise Drive
Westerville
Ohio 43081
U.S.A.

Isobel Clark
Geostokos Ltd.
P.O. Box 1897
London
NW6 lDS
England

ABSTRACT. The geostatistical techniques of conditional simulation are well documented for the univariate case. This paper presents a case study in geostatistical simulation based on the estimation of multivariate data. The potentiometric pressure of groundwater within an aquifer is estimated using borehole data from that aquifer and also from a neighbouring (correlated) aquifer. The sample data used is public domain and comes from the Texas Panhandle region in the United States. The case study illustrates a reasonably simplistic approach to the multivariate conditional simulation and discusses the advantages of the *Co-Kriging* method over the univariate approach. The use of sparse sample data provides motivation for using the newer variation on co-kriging, generally referred to as M.U.C.K.

Introduction

Deaf Smith County in northern Texas was, at one time, a proposed site for a high level nuclear waste repository. The planned site for the storage of this waste was within a salt bed, which is underlain by two aquiferous strata known as the Wolfcamp and the Pennsylvanian. The general geological characteristics of these aquifers - and the borehole data used in this study - are published elsewhere (cf. Harper & Furr; Harper, Basinger & Furr) and will not be discussed in detail here.

One of the essential factors to characterise in the assessment of the repository is the risk of hazardous material being released into one or other of these aquifers and, hence, transported to potentially remote sites. Amongst other parameters, the pressure of fluid within the aquifer will determine the 'path' taking by a freed particle and its potential final destination. The potentiometric level can be estimated over the study area and maps of predicted pressure produced. From this we can assess the likely travel path of a particle freed at any specified site. However, this estimated surface will be smoother than the 'real' potentiometric variation — adding another source of uncertainty to the predicted travel path of the radio-nucleid. The missing 'roughness' can be reproduced by simulation of the estimation errors around the interpolated surface. For this purpose we can use standard geostatistical conditional simulation techniques, such as that known as "turning bands".

The major feature of the case study discussed in this paper is that borehole data is scarce in the regions to which the particles are likely to travel. This is particularly true of the Pennsylvanian aquifer, for which little data is available in the north-east of the study area. We discuss the improvement of the basic estimation through a multivariate co-kriging approach and the multivariate simulation required to reproduce the roughness on a surface estimated by co-kriging. The impact of this approach on the predicted travel paths of a freed radio-nucleid are illustrated.

Potentiometric Pressure within the Pennsylvanian

ESTIMATION OVER THE STUDY AREA

In this paper we discuss the estimation of the travel path of a freed radio-nucleid within the Pennsylvanian aquifer. 109 boreholes intersecting the aquifer are available for the analysis. The surface has a strong polynomial-type trend, which must be removed before the semi-variogram is constructed and a model fitted. The geostatistical analysis of this sample data has been discussed elsewhere (cf. Clark, Basinger & Harper). The model fitted to the Pennsylvanian borehole data comprises a local linear trend surface, a spherical semi-variogram component with a range of influence of 50 miles and a sill of 46,000 ft^2, plus a nugget effect of 12,000 ft^2. This model was found to cross validate at an acceptable level.

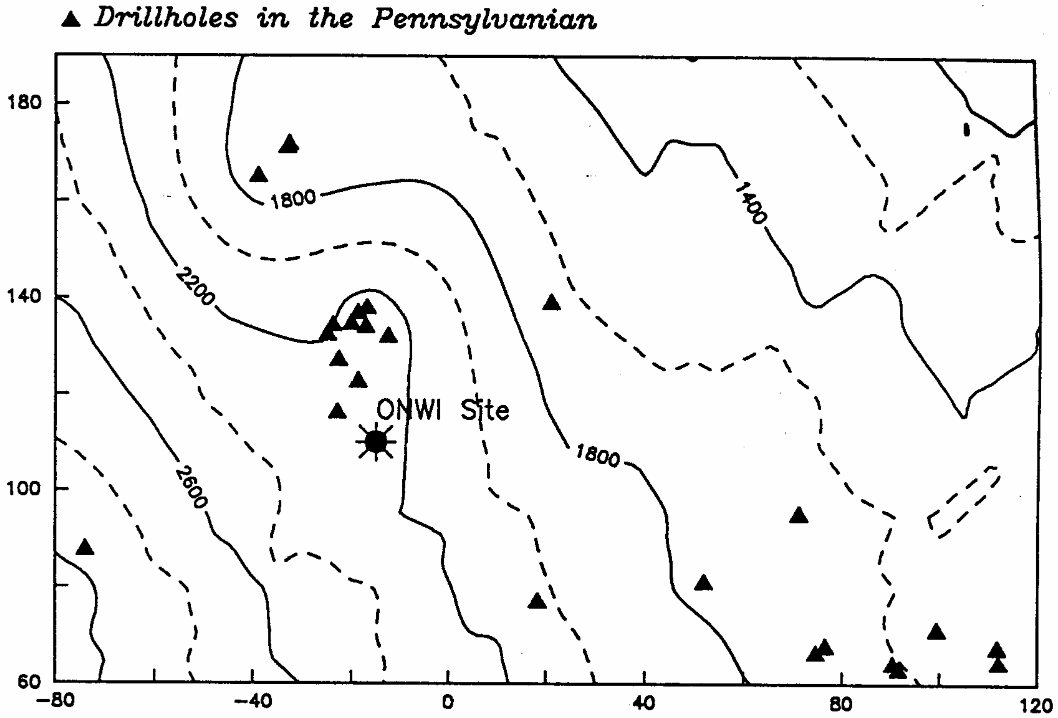
Using this semi-variogram analysis, the technique known as Universal Kriging can be used to interpolate values onto a grid of points. This grid may then be used to produce contour maps. Figure 1 shows the potentiometric surface over the area local to the proposed repository, rather than the whole study area. These contours have been produced by estimating values at the nodes of a 5 mile grid. It can be seen that the surface is apparently very smooth, leading us-to suppose that we can predict the travel path of a freed particle with ease. However, for every estimated grid point there is an associated 'standard error' which quantifies the uncertainty on the estimate at that particular location. The corresponding map of standard errors on this Universal Kriging exercise is shown as Figure 2 overleaf.

PREDICTING THE TRAVEL PATH

Let us assume that a radio-nucleid is released at a potential repository site, shown on the maps as "ONWI Site". It is assumed that the particle will travel from a grid point of higher pressure to one of lower potentiometric level. In many cases it has been assumed that the particle will always take the 'steepest' path. In this study, we have simply assumed that any neighbouring grid point which has a lower potentiometric level is a likely destination for the particle. It is hoped in later studies to add a preferential factor for steeper slopes - without reverting to the single path approach.

Assuming that the particle is freed at the ONWI Site, we can trace all the potential paths which might be taken. In this fashion a fan of travel paths and destinations may be evaluated. We display this fan in a map such as Figure 3. The contours on this map represent the number of ways a grid point can be reached by a particle starting out from the ONWI site. Thus, a value of '4' implies that this particular grid point may be reached from four neighbouring points - each of which may be reached (in some way) from the ONWI site. The maximum number of ways each grid point may be reached is 8, since we allow diagonal movement between grid points.

Figure 1: Potentiometric levels estimated by Universal Kriging



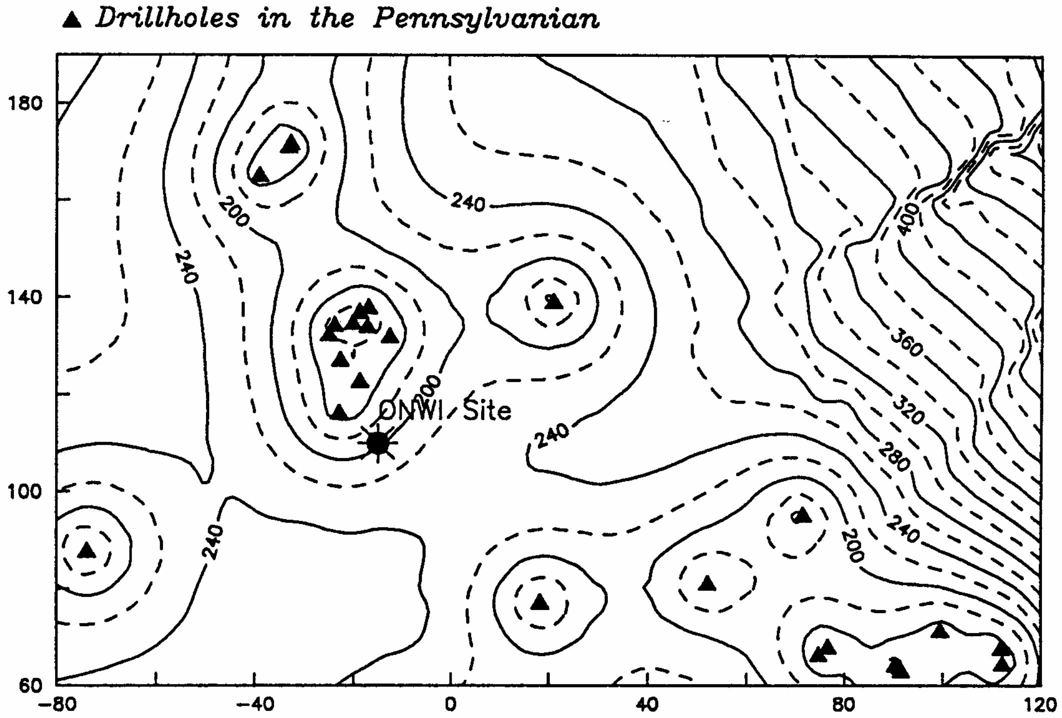


Figure 2: Standard errors on UK'ed potentiometric levels

In this fashion, we may outline the predicted risk of a freed particle arriving at a particular location within the study area. However, this is a *best estimate* of the travel fan. The true potentiometric surface will be 'rougher' than the kriged surface - as evidenced by the nugget effect present in the semi-variogram model. To obtain some idea of how the true travel path might differ from this best estimate we must turn to simulation.

CONDITIONAL SIMULATION

The geostatistical technique known as conditional *simulation* is very widely used, having been introduced and discussed in many papers and text books for over fifteen years (cf. Journel & Huijbregts). As a brief summary of the method:

- A geostatistical model is fitted to the study area, in the form of semi-variogram models and (possibly) a trend component;
- A statistical distribution is established for the sample values (or for the residuals from any trend);
- A Kriging method is used to estimate the values at grid points over the study area - and their associated standard errors;
- An independent surface is simulated, having the same semi-variogram and statistical distribution as the original values;
- Values at the sampled locations and at the grid points are extracted from the simulated surface;
- Using the simulated samples, the kriging exercise is repeated - resulting in a simulated kriged surface.

The above process results in three grids: the original kriged values (smoothed), a simulated grid (rough) and a simulated kriged grid (smoothed). If the procedure is carried out correctly, the difference between the simulated grid and the simulated kriged grid should emulate the difference between the 'true' surface and the actual kriged grid. A simple sum will, therefore, produce the roughness which must then be 'added' to the kriged surface to produce a simulation which conforms to the original data but possesses the sort of variability expected from a real surface.

Of course, we do not carry out one simulation but as many as possible. Figure 4 shows the results obtained after 100 simulations of the above type. As in Figure 3, the contours represent the number of ways in which a particular grid point can be reached from the ONWI Site starting point. However, in this case the contours are the summation of 100 simulations rather than a single case. The most notable features of the simulated fan are twofold: (1) the particle can travel in directions which were estimated to be 'uphill'of the kriged surface; and (2) there are locations which the kriged surface predicted as 'reachable' which have not been reached by any one of our 100 freed particles. This latter point does not imply that they would not have been reached on other simulated surfaces. The former point illustrates the variability which actually exists around our best estimated surface. Although a particular grid point may be *estimated* as having a higher potentiometric level, the uncertainty on this estimate shows the possibility that it might really have a lower level than the repository site.

A Multivariate Approach

USING DATA FROM THE WOLFCAMP

It may be asked: if we can produce these results from a univariate approach, why is there a need for a multivariate approach. Simply this, the borehole data available from the Pennsylvanian itself is sparse - not so much in the region of the ONWI site, but over that area into which the particle would be expected to travel. Better quantification of the uncertainties in this area will lead to better estimation, better simulation and, hence, better assessment of the attendant risks involved in the release of hazardous material.

Figure 5 overleaf shows the locations of boreholes in the Wolfcamp aquifer (stars) as well as those in the Pennsylvanian (triangles). The Wolfcamp has been modelled as having a linear trend (locally), a spherical semi-variogram component with a range of influence of 60 miles and a sill of 34,000 ft^2, plus a nugget effect of 11,000 ft^2 (op cit). This model would be sufficient if we wished to estimate values within the Wolfcamp aquifer from samples within that aquifer. To use the Wolfcamp information to assist in the estimate of the Pennsylvanian, we must assess the relationship between Wolfcamp and Pennsylvanian potentiometric levels. In a geostatistical context, this implies that we need to construct a cross-semi-variogram between the Wolfcamp values and the Pennsylvanian values. In 'classical' co-kriging, such a semi-variogram is constructed as follows:

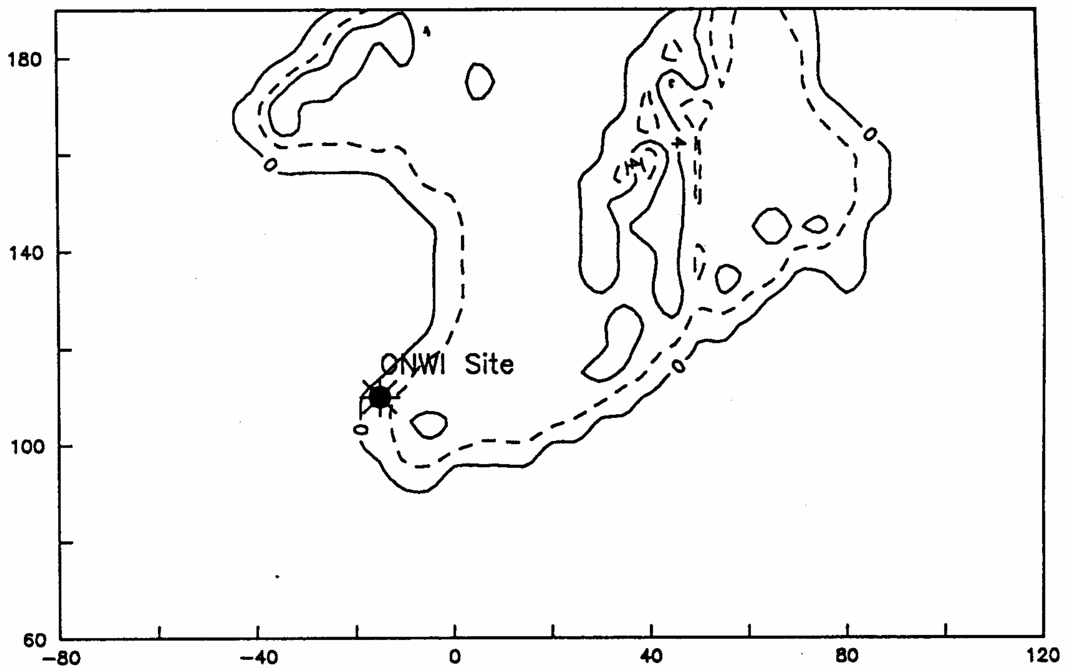
$$\gamma_{gf}^*(h) = \Sigma(g_i - g_j)(f_i - f_j)$$

where g_i denotes the ith observation (say) in the Wolfcamp and f_j the jth observation in the Pennsylvanian. This is a direct analogy with the univariate definition:

$$\gamma^*(h) = \Sigma(g_i - g_j)^2$$

and also converts easily into the covariance form under stationarity (d. Carr & McCallister; Myers). This approach has one major drawback in our current application: it requires that we have both measurements made at the same locations, i.e. we must have both g and f measured at the same sites. As can be seen from Figure 5, this is a rather unrealistic condition to apply to our case study. Within the whole study area - rather than the locality of the ONWI Site - we have only eight or so locations at which both potentiometric levels have been measured. This is not sufficient to characterise a semi-variogram model.

Figure 3: Travel Path Fan for UK'ed potentiometric levels



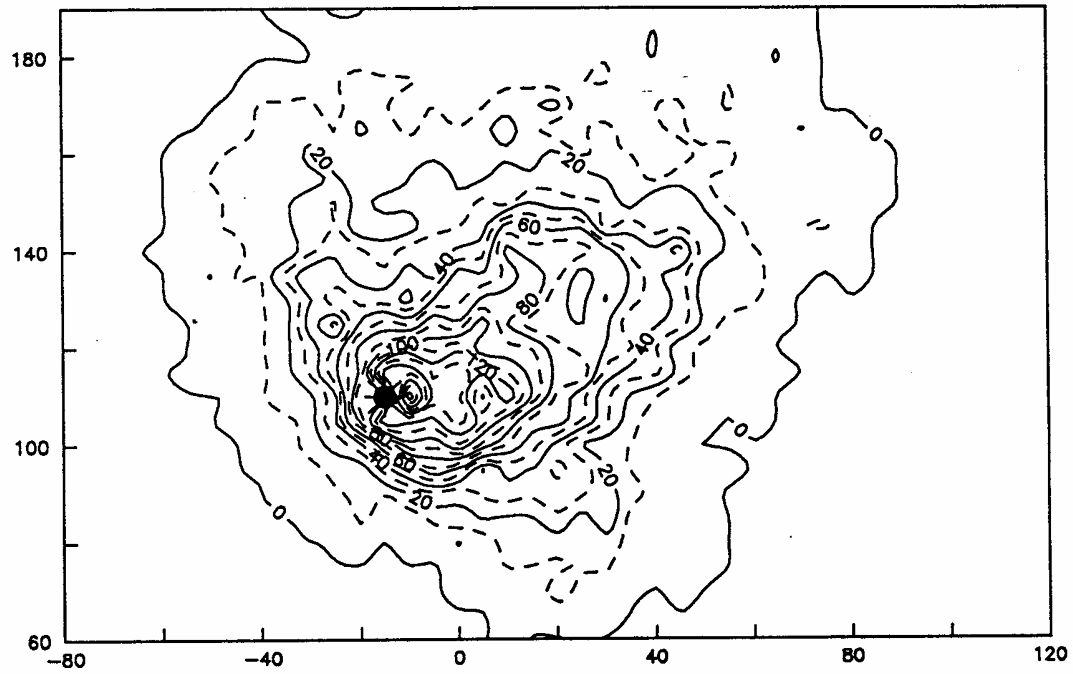


Figure 4: 100 Simulations based on UK solution

AN ALTERNATIVE FORMULATION

In situations where the separate variables have been measured at different locations, an alternate form for the semi-variogram has been proposed (op *cit):*

$$\gamma_{gf}{}^*(h) = \Sigma(g_i - f_j)^2$$

This is also analogous to the univariate case, but does not simplify to a covariance form. Possible limitations on this form of cross-semi-variogram exist if the two variables have markedly different mean values or widely different variances. The latter is true of any cross-covariance approach, of course. In our case study, we have subtracted a trend from each of the aquifers separately, before attempting to calculate semi-variograms. Removing the trend leaves residuals with a zero mean, so that the former problem should not trouble us in this case study.

A cross-semi-variogram between the Wolfcamp residuals and the Pennsylvanian residuals may be modelled satisfactorily by a spherical semi-variogram with a range of 35 miles and a sill of 28,000 ft^2, plus a nugget effect of 15,000 ft^2. Cross validation exercises may be carried out in a similar way to the univariate case.

KRIGING ESTIMATES

To distinguish kriging results obtained by this type of semi-variogram modelling, we refer to the estimation technique as "multivariate universal co-kriging", rather than simply co-kriging - which we take to imply the classical approach. The contours on Figure 5 were obtained by M.U.C.K. and Figure 6 shows the corresponding uncertainty in the estimates, in the form of a standard error map.

The improvement in estimation afforded by the inclusion of the Wolfcamp data is immediately apparent.

SIMULATION FROM MUCK RESULTS

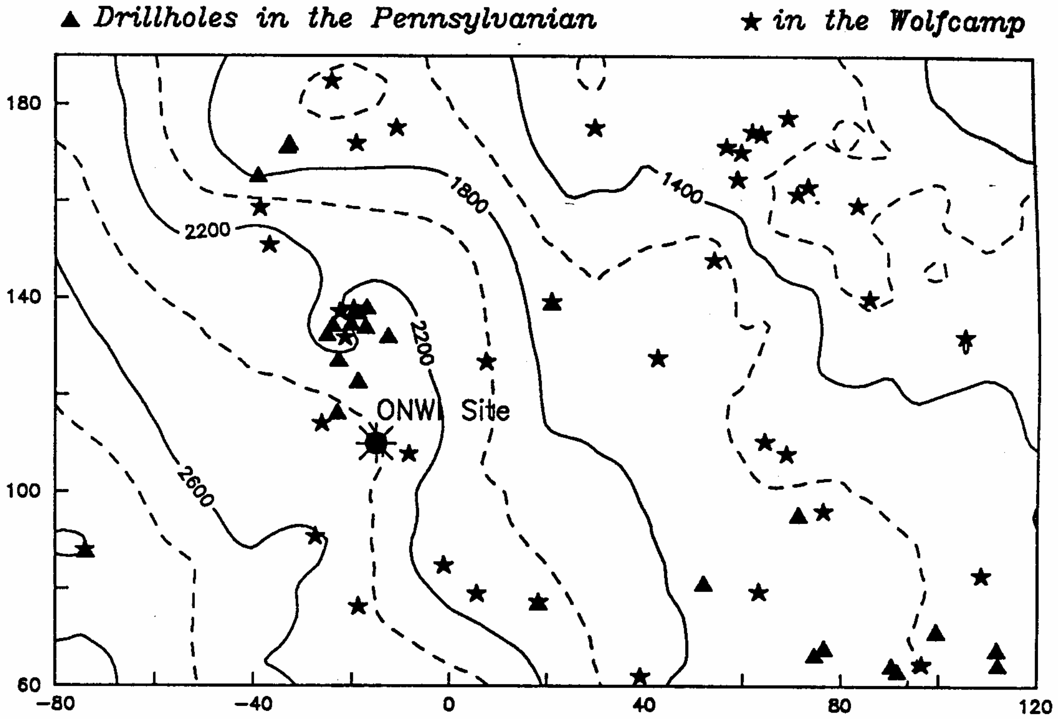
The multivariate conditional simulation is probably best explained by referring to the series of steps listed above in the univariate section:

- A geostatistical model is fitted to the study area, in the form of semi-variogram models and (possibly) a trend component;

In the multivariate case, this stage is extended by modelling each variable separately, and each cross-semi-variogram between pairs of variables.

- A statistical distribution is established for the sample values (or for the residuals from any trend);

Figure 5: Potentiometric levels estimated by MUCK



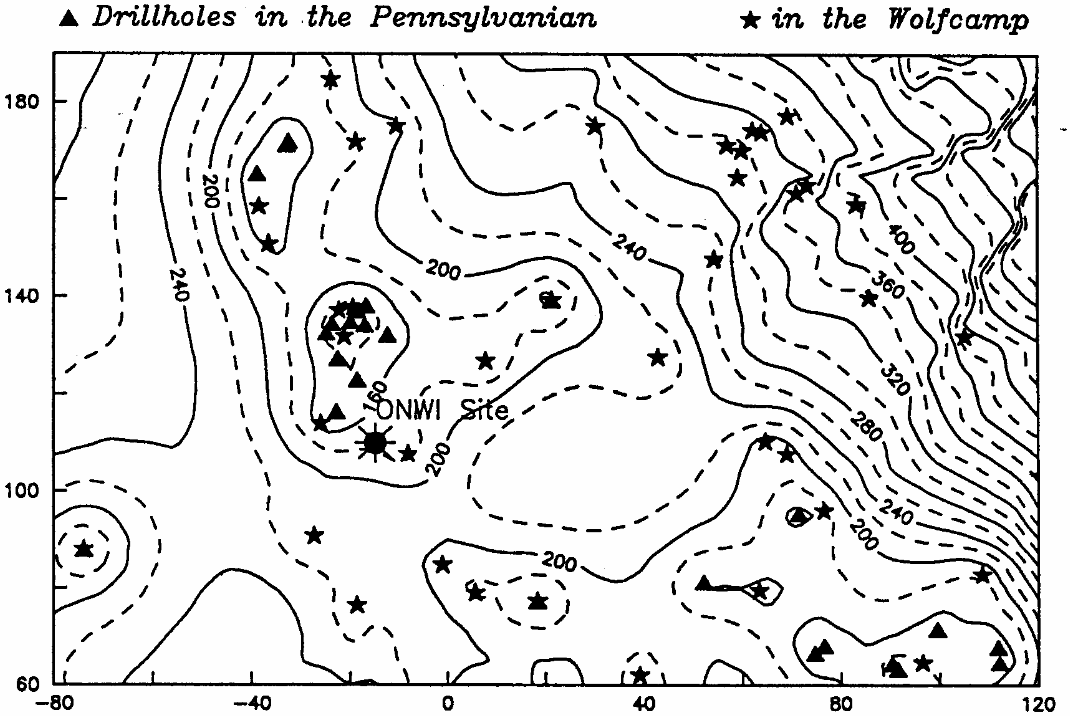


Figure 6: Standard errors on MUCK'ed potentiometric levels

For several variables, we must establish a vector of means, a variance/covariance matrix and multivariate Normality (or a transform to it)-

- A Kriging method is used to estimate the values at grid points over the study area - and their associated standard errors;

Using M.U.C.K. (say) rather than Ordinary or Universal Kriging.

- An independent surface is simulated, having the same semi-variograrn and statistical distribution as the original values;

This stage will be discussed in detail later in this paper.

- Values at the sampled locations and at the grid points are extracted from the simulated surface;

Using the simulated samples, the kriging exercise is repeated - resulting in a simulated kriged surface.

These stages remain the same, with the proviso that co-kriging is used in the last stage. In fact, the only tricky stage in the generalisation to the multi-variable case is in the basic unconditional simulation of multivariate data with the correct means, variance/covariance matrix, semi-variograms and cross-semi-variograms. We have adopted a fairly simplistic approach to this problem which yields more than adequate cross validation analyses. To explain this approach, it is necessary to examine the geostatistical 'unconditional' simulation stage in a little more detail. When using the turning bands approach, the simulations are actually unidirectional. That is, values are simulated at regular intervals along -a line - then the lines are combined in such a way as to produce a three-dimensional simulation which conforms to the required semi-variogram. The unidirectional simulations are produced in two stages:

1. Independent random values are simulated at regular intervals along the line;

2. A weighted moving average is applied which results in the correct pseudo-semi-variogram along the line.

A full mathematical solution to the multivariate analogue of this process is feasible. However, we have found that a simplified approach yields acceptable (and verifiable) results. The method proposed is as follows:

1. Independent random vectors are simulated from the required multivariate Normal distribution, with specified means and suitably modified variance/covariance matrix, at regular intervals along the line;

2. each variable is then subjected separately to the weighted moving average relevant to its *univariate* semi-variogram model.

All cross validation techniques that we have been able to study show this method to reproduce the required sort of variability for conditional simulation of the multivariate situation.

The Final Product

Using the co-kriging estimation technique and the multivariate conditional simulation, it is possible to produce travel path fans from a model with less associated uncertainty. By incorporating the Wolfcamp information into the estimation of values in the Pennsylvanian, we can better characterise the likely destination of a freed radio-nucleid. 100 simulations were carried out, to give a direct comparison with the univariate case described earlier in the paper. Figure 7 shows the travel path fan from these 100 simulations. Comparing this with Figure 4, we can see that the fan is wider - that is, our univariate simulation was still (in some respects) over-optimistic about the likely destinations of the particle. The northeast corner of the study area, in particular, can be reached by a freed particle although the probability of this happening is fairly low.

In a full study, of course, many more than 100 simulations would be carried out and a wider area studied. However, the basic problems of sparse data would remain. Multivariate conditional simulation seems to provide one reliable method of assessing the uncertainty in the travel path of hazardous waste if released from the proposed repository.

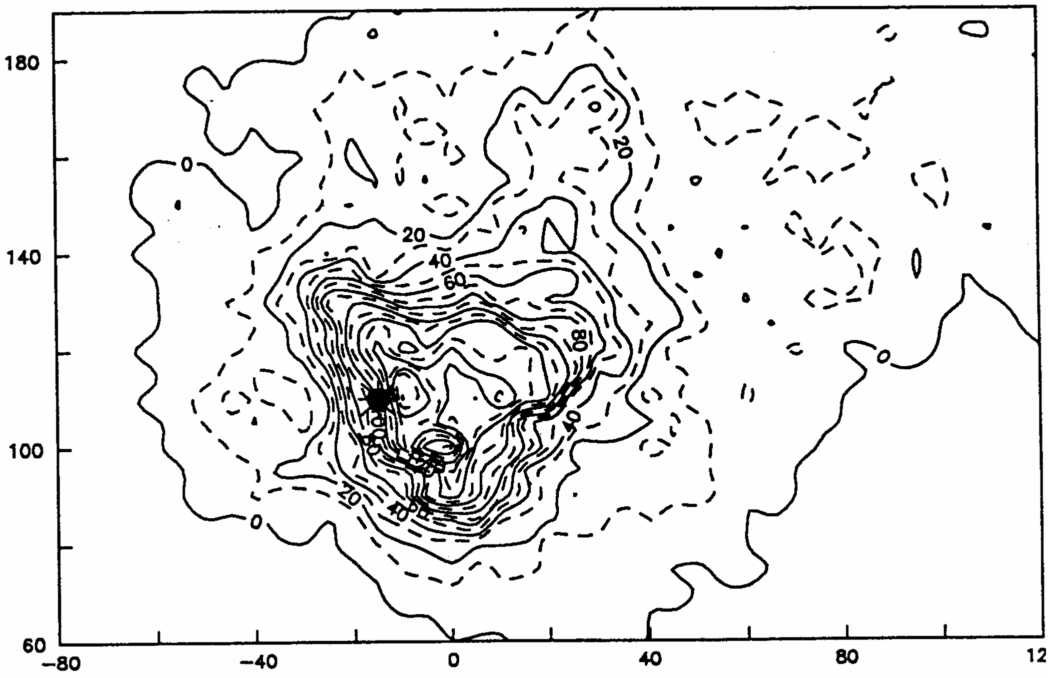


Figure 7: 100 simulations based on MUCK solution

References

Carr, J.R. & McCallister, P.G. (1985) "An application of co-kriging for estimation of tripartite earthquake response spectra", *Mathematical Geology,* 17, 5, 527-545.

Clark, I., Basinger, K.L. & Harper, W.V. (1987) "MUCK - a novel approach to cokriging", *Proc. AECL/DOE meeting on Geostatistics and SensitivitylUncertainty Analysis,* San Francisco, Battelle Press, 473-493.

Harper, W.V. & Furr, J.M. (1986) "Geostatistical analysis of potentiometric data in the Wolfcamp aquifer of the Palo Duro Basin, Texas", *BMI/ONWI-587,* Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, Ohio.

Harper, W.V., Basinger, K.V. & Furr, J.M. (1986) " Geostatistical analysis of potentiometric data in the Pennsylvanian aquifer of the Palo Duro Basin, Texas", *BMI/OIVWI-680,* Office of Nuclear Waste Isolation, Battelle Memorial Institute, Columbus, Ohio.

Journel, A.G. & Huijbregts Ch.J. (1978) Mining Geostatistics, Academic Press, New York.

Myers, D.E. (1984) "Co-Kriging - new developments", in *Geostatistics for Natural Resources Characterization,* G. Verly et al (Eds), D. Reidel, 295-305.