Stochastic simulation for the quantification of mine spoil variability and rehabilitation decision making

R. Dimitrakopoulos and S. Mackie

WH Bryan Mining Geology Research Centre, The University of Queensland, Brisbane, Qld 4072

s.mackie@mailbox.uq.edu.au

Abstract: Soil attributes, including those in mine spoil heaps, critically affect plant growth during land rehabilitation. Their characterization through a limited number of samples requires quantification of spatial variability, which is then used at various stages throughout the rehabilitation process and assists risk analysis and rehabilitation decision making. Stochastic simulation is a tool used for the quantification of uncertainty. This paper presents the suitability of stochastic simulation for the joint simulation of soil attributes and introduces a new computationally efficient method. The method is based on: (i) the Minimum/Maximum Autocorrelation Factors (MAF), involving the de-correlation of pertinent variables into spatially non-correlated factors, and (ii) the simulation of MAF and back transformation to the conditional simulations of the correlated variables. MAF factors in point (ii) are simulated using the new Generalised Sequential Gaussian Simulation (GSGS) technique which is substantially more efficient that the traditional sequential simulation methods. The formulated approach is applied to mine spoil data, specifically electrical conductivity and pH, which typically contribute to restricted plant growth on spoils in Queensland coal mines. The results of the simulations are used to quantify risk of exceeding significant thresholds for each variable, thereby identifying problem rehabilitation areas. The case study demonstrates the practical aspects of the method as well as its use in planning rehabilitation strategies and predictions of future performance of the rehabilitation.

Keywords: *Minimum/maximum autocorrelation factors; joint simulations; generalised sequential Gaussian simulation, land rehabilitation.*

1. INTRODUCTION

Soil properties critically influence the decision making involved in the development, implementation and monitoring of rehabilitation programs. As soil properties are only sampled at a limited number of locations, the quantitative characterization of their distribution and variability at any unsampled location is critical. The assessment of local uncertainty about possible values is a well established issue in soil science and environmental engineering (Webster and Oliver, 1989; Pachepsky and Acock, 1998; Bross et al., 1999) and stochastic simulation is a key tool used in modeling this uncertainty (e.g. Goovaerts, 2001). However, these issues have received little attention to date in relation to the rehabilitation of mined land and related concerns for plant establishment, survival and long term sustainability. Environmental impact mitigation and concurrent reclamation are now commonly regarded as integral components of the mining process (Morrey, 1999). Thus, there is an increased need to develop suitable modelling frameworks for the prediction of environmental changes, and hence the assessment of impacts.

A key issue in modelling soil properties, including mine spoil heaps, is that the modelling of several commonly correlated properties of soils are needed. Properties including alkalinity, electrical conductivity, salinity, sodium content, nitrate concentrations and phosphates that may affect plant success (Grigg et al., 2000) show spatial cross-correlations. Techniques to jointly simulate spatial distributions of soil attributes are available (e.g. Gutjahr et al., 1997) and improve the plausibility of resultant models. However, they are computationally intensive. Contributors to complexity include the tedious inference and modelling of cross-correlations and computational inefficiencies, both substantially increasing with the number of variables being jointly-simulated. A practical alternative to the ‘direct’ joint-simulation of variables is the decorrelation of variables introduced using principal component analysis or PCA (David, 1988; Wackernagel, 1995). The effectiveness of this approach, in the presence of spatial cross-correlations, is limited because PCA does not eliminate cross-correlations at distances other than zero. To overcome the above limitations, minimum/maximum autocorrelation factors, MAF, (e.g. Desbarats and Dimitrakopoulos, 2000) are used to de-correlate pertinent variables into spatially non-correlated factors that are independently simulated and back transformed to correlated attributes. The simulations of MAF are generated with the new fast generalised sequential Gaussian simulation, GSGS, (Dimitrakopoulos and Luo, 2003) to provide a substantially more efficient simulation framework.

The following sections will firstly introduce the method of joint simulation of multiple correlated variables based on MAF. A description of the data available follows together with the results of the joint simulation. An application of risk analysis is then presented towards assisting with rehabilitation strategies followed by conclusions.

2. JOINT SIMULATION OF CORRELATED VARIABLES WITH MINIMUM/MAXIMUM AUTOCORRELATION FACTORS

In geostatistical terminology, the attributes of elements in soils are represented by a multivariate stationary and ergodic random function. Consider a multivariate, ℓ dimensional, Gaussian, stationary and ergodic spatial random function $\mathbf{Z}(\mathrm{x}) = [\mathrm{Z}_1(\mathrm{x}),...,\mathrm{Z}_\ell(\mathrm{x})]^T$. Minimum/Maximum Autocorrelations Factors are defined as the ℓ orthogonal linear combinations $Y_i(\mathrm{x}) = a_i^T \mathrm{Z}(\mathrm{x}),\ i = 1,...,\ell$ of the original multivariate vector $\mathbf{Z}(\mathrm{x})$. MAF are derived assuming that $\mathbf{Z}(\mathrm{x})$ is represented by a two-structure linear model of coregionalisation (Wackernagel, 1995). The MAF transformation can be rewritten as

$$\mathbf{Y}(x) = \mathbf{A}_{\mathrm{MAF}}\, \mathbf{Z}(x) \tag{1}$$

and the MAF factors are derived from

$$\mathbf{A}_{\mathrm{MAF}} = \mathbf{Q}_2 \mathbf{\Lambda}_1^{-1} \mathbf{Q}_1 \tag{2}$$

where the eigenvectors $\mathbf{Q}_1$ and eigenvalues $\mathbf{\Lambda}_1$ are obtained from the spectral decomposition of the multivariate covariance matrix $\mathbf{B}$ of $\mathbf{Z}(\mathrm{x})$ at zero lag distance. More specifically,

$$\mathbf{Q}_1 \mathbf{B} \mathbf{Q}_1^T = \mathbf{\Lambda}_1 \tag{3}$$

and $\mathbf{Q}_2$ is the matrix of eigenvectors from the spectral decomposition

$$\mathbf{Q}_2 \mathbf{M}(\Delta) \mathbf{Q}_2^T = \mathbf{Q}_2 \left(\frac{1}{2} \left[[\mathbf{\Gamma}_Y(\Delta)]^T + [\mathbf{\Gamma}_Y(\Delta)] \right] \right) \mathbf{Q}_2^T \tag{4}$$

where the matrix $\mathbf{\Gamma}_Y(\Delta)$ is an asymmetric matrix variogram at lag distance Δ for the regular PCA factors $\mathbf{Y}(x)$=$\mathbf{Z}(x)$ $\mathbf{A}$, where $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}^{-1/2}$. In practice, several Δ lag distances may be used for values lower than the range and the resulting eigenvectors averaged.

Given the MAF transformation above, the joint simulation of multiple correlated variables using the MAF approach proceeds as follows:

i. Normalize the variables to be simulated.
ii. Use MAF to generate the MAF non-correlated factors.
iii. Produce variograms for each MAF.
iv. Conditionally simulate each MAF using a Gaussian simulation method.
v. Validate the simulation of factors.
vi. Back-transform simulated MAF to variables and denormalize.
vii. Validate the final results.
viii. Generate additional simulations, as needed.

The conditional simulation of a Gaussian random function Y(x) above is based herein on the decomposition of the multivariate probability density function (PDF) of a stationary and ergodic random function to a product of local conditional distributions

$$f\left(x_1,...,x_N; y_1,...,y_N\right) = \prod_{i=1}^{N} f\left(x_i; y_i \mid (n+i-1)\right) \tag{5}$$

where $f\left(x_1,...,x_N; y_1,...,y_N\right)$ is the pdf, N the number of points discretising the field to be simulated, n the number of available data, and x_i the location of a point in the space considered.

Setting $\Lambda_i = \Lambda_0 + y\left(x_\alpha\right)$, α=1,...,i , where $y\left(x_\alpha\right)$ is a realisation of Y(x) at location x_α, and considering groups of N_p nodes, Eq. (5) is

$$f\left(x_1,...,x_N; y_1,...,y_N \mid \Lambda_0\right) = \prod_{i=1}^{N_1} f\left(x_i; y_i \mid \Lambda_{i-1}\right) \bullet \prod_{i=N_1+1}^{N_1+N_p=N_2} f\left(x_i; y_i \mid \Lambda_{i-1}\right) \bullet ....\bullet \prod_{i=N_{k-1}+1}^{N_{k-1}+N_p=N_k} f\left(x_i; y_i \mid \Lambda_{i-1}\right) \tag{6}$$

The simulated group of nodes in vector $\mathbf{y}_p$ is then

$$\mathbf{y}_p = \mathbf{C}_{pI} \mathbf{C}_{II}^{-1} \mathbf{y}_I + \mathbf{L}_{pp} \mathbf{w}_p \tag{7}$$

where the covariance matrix $\mathbf{C}_{II}$ is the covariance between the conditioning nodes, $\mathbf{C}_{pI}$ the covariance between the group of nodes to be simulated and the conditioning nodes, and $\mathbf{C}_{pp}$ the covariance between the nodes of the group; $\mathbf{y}_I$ is the data vector contained in Λ_{i-1} and $\mathbf{w}_p$ a standard normal random vector. The generalised sequential Gaussian simulation algorithm (GSGS) from Eq. 6, used to simulate the N nodes in a domain D is:

i. Define a random path visiting each group of N_p nodes to be simulated.
ii. At each group of nodes, use Eq. (7) to generate simulated values and add the values to the data set.
iii. Go to the next group of nodes and repeat the previous two steps.
iv. Loop until all groups of nodes have been visited and the N nodes simulated.

3. THE DATA AVAILABLE

The data used in this study were sampled from a mined waste (spoil) dump of an open cut coal mine in Queensland. They include measurements of electrical conductivity (EC), a measure of salinity, and pH, a measure of acidity.

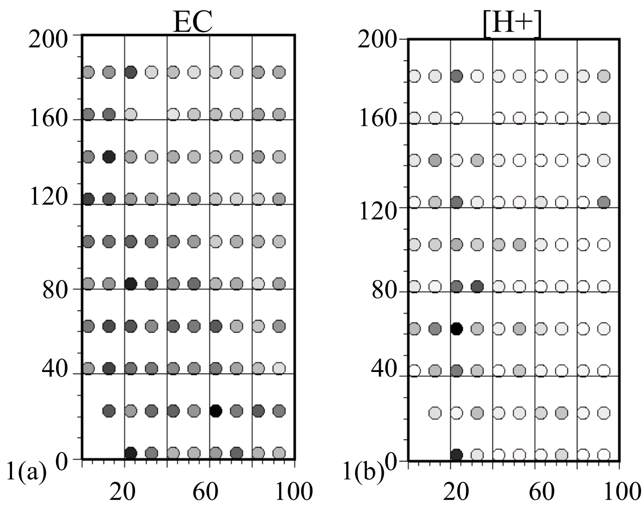


Figure 1. Measured properties across waste dump at a Queensland coal mine. (a)Electrical conductivity and (b) acidity.

The dataset is composed of 96 sampled locations on a regular grid with EC and pH measurements in all locations. pH values are converted to concentration of hydrogen ions in solution, [H+]. These data are characteristic of the conditions found in spoil dumps of many Queensland coal mines in that they are highly alkaline and highly saline. Salinity may be a major impediment to plant establishment and survival, while high levels of pH limit the availability of phosphorus, which in turn limits plant growth and sustainability (Grigg et al., 2000). The spatial distribution of both variables is shown in Figure 1. Dark circles represent high values and light circles represent low values. The heterogeneous nature of the spoil is clearly evident.

4. JOINT SIMULATION OF SPOIL PARAMETERS: EC AND [H+]

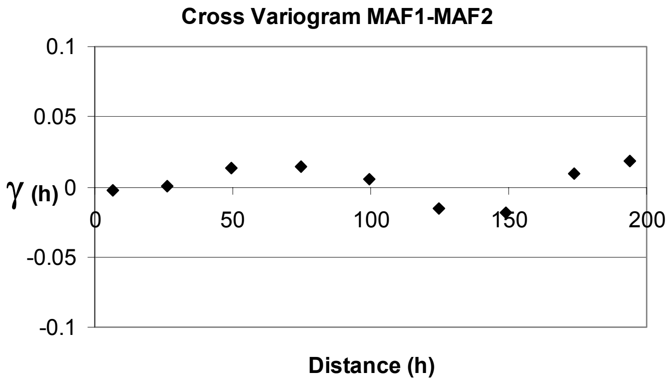
4.1. Normal-score transformation

Following the simulation steps using MAF described earlier, a normal-score transformation is performed on the distribution of EC and [H+] data. Normal score transformations are based on rank ordering of the data and decrease the influence of outliers. This, in turn, assists the inference of the variogram and estimation of covariance matrices in the simulation process that follows.

4.2. MAF transformation

The transformation matrix $\mathbf{A}_{\mathrm{MAF}}$ (Eq. 1) used to generate the min/max autocorrelation factors is shown in Figure 2 (b). MAF are calculated by multiplying the vector of EC and [H+] by a vector of loadings from the rows of the transformation matrix. It should be noted that the MAF loadings are quite different from the ones derived by PCA (Desbarats and Dimitrakopoulos, 2000). The lag Δ in Eq. 4 used in this example is 65 metres and was derived experimentally by testing several lag distances to assure a suitable decorrelation and stable MAF decomposition. Figure 2 (a) shows the cross-variogram between MAF from the present study and demonstrates variable decorrelation. Experimental variograms and cross-variograms for EC and [H+] are shown and discussed in more detail in a subsequent section.

2(a)



0.756	0.001
0.654	0.999

2(b)

Figure 2. (a) Cross-variograms of MAF and (b) transformation matrix.

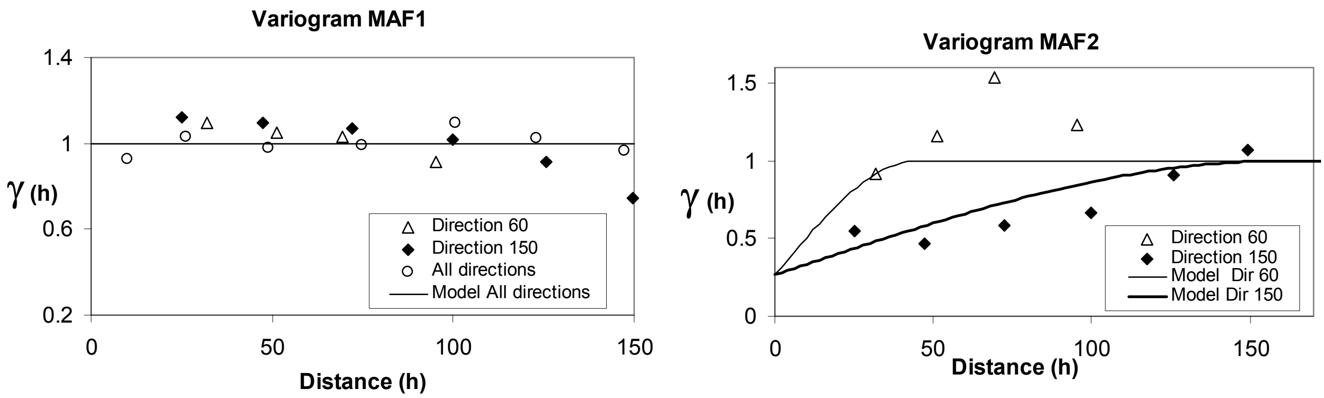


Figure 3. Experimental and model variograms of MAF.

4.3. Variography of MAF

Variography on each MAF is performed. Figure 3 shows the experimental and model variograms fitted to the both MAF. Note that variogram models for MAF2 are spherical and show clear spatial patterns. MAF1 is modeled as pure nugget. MAF variograms are subsequently used in the simulation of each factor and the validation of the MAF simulation results. It should be noted that MAF variograms are linear combinations of the variograms of the original (normal score) variables.

4.4. Conditional simulation of MAF

Conditional simulation is performed independently on both MAF using the GSGS algorithm. The simulations are performed on a grid of 5000 nodes within the limits of the waste dump. Thirty simulations are generated in this study and are validated in detail for reproduction of data, histograms and variograms. The validation of the MAF simulations is not presented here as a subsequent section presents the validation of realizations in the data space.

4.5. Back transformations of MAF

The realizations of MAF were transformed back to simulated normal score variables by multiplying a column vector of simulated MAF in each grid node with the corresponding inverse matrix of the MAF loadings in Figure 2 (b). Subsequently, the normal score EC and [H+] realizations are back transformed to the data space.

4.6. Validation of the joint EC and [H+] simulation results

Several validation checks are performed to assess the results of the joint simulations of EC and [H+] using the MAF transformations. Validation involves calculation of histograms, experimental variograms and cross-variograms of the simulated realizations to ensure reproduction of original data and their spatial characteristics.

Figure 4 shows plots of variograms and cross-variograms for the original data and conditional simulations. All results suggest that the reproduction of the original data spatial characteristics by the simulated realizations is excellent. Recall that the variograms and cross-variograms of original variables are not directly

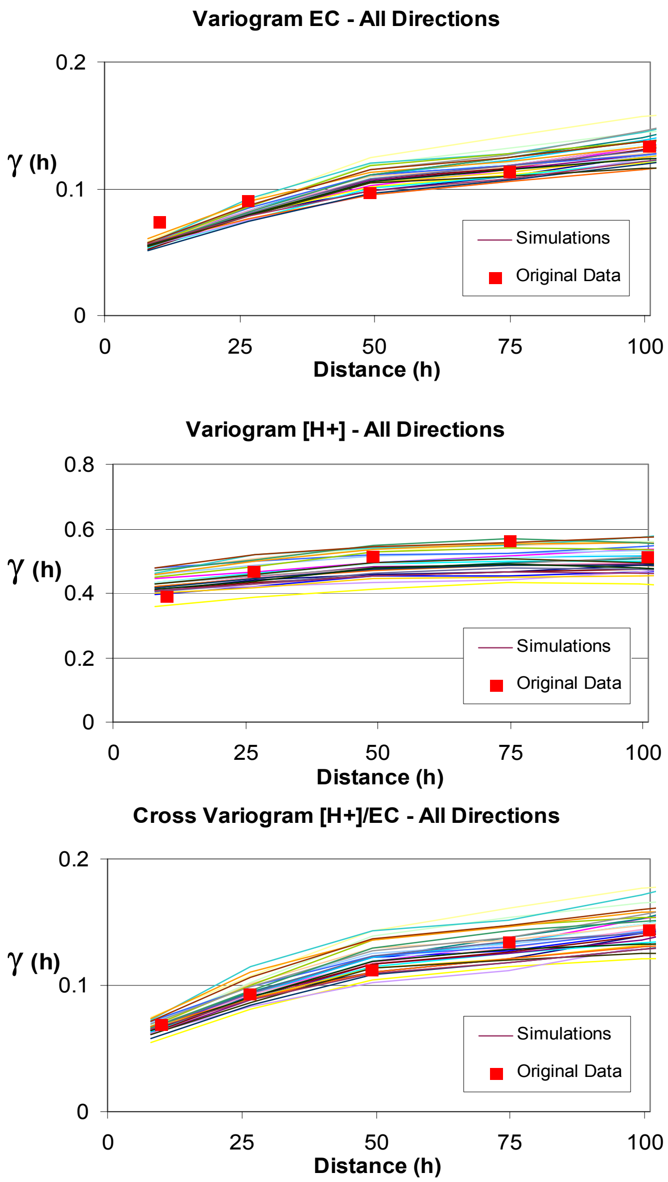


Figure 4. Variograms and cross-variograms of simulations backtransformed to the data space, compared to experimental variograms of original data.

used in the joint simulation based on MAF, which used the variograms of the independent MAF.

5. RISK ANALYSIS FOR DECISION MAKING

The main objectives of mined land rehabilitation include a sustainable land use after mining, stability of the land surface, and preservation of water quality (Grigg et al., 2000). Currently, there are no formal criteria used to assess the success of rehabilitated areas for mines. For open cut coal mines in Queensland, such as the case study in the previous section, suitable completion criteria for relinquishment purposes and, more specifically, pasture–based rehabilitation in Queensland have been established (Grigg et al., 2001). These criteria suggest the achievement and maintenance of at least 70% vegetation cover because there is considerable evidence, from minesite erosion research and elsewhere, that vegetation cover is the single most important factor affecting soil loss in rehabilitated pastures.

Analysis of data from monitoring studies upon which rehabilitation criteria were based, indicate a strong influence of average salinity (EC) on pasture performance, indicated by the amount of total dry matter (DM) (Grigg et al., 2001). An exponential improvement in DM is evident for decreasing salinity levels. The amount of DM is also related to ground cover and an exponential increase in ground cover is evident with increasing DM. These relationships indicate that an EC of 0.6 would permit development of sufficient dry matter to achieve ground cover of 70% (Grigg et al., 2001).

We have shown that it is possible to simulate spoil parameters throughout the waste dump and that these simulations enable us to quantify the variability of certain parameters. These simulations can then be used to assess the probability, or risk, that EC will exceed 0.6, the cut off value as discussed above to ensure 70% vegetation cover. This is determined from the number of simulations in which the generated value for a given location is above the cut off value. In this example, it was determined that a cover of 70% was required for successful rehabilitation. Other goals may be tested, however, by assessing different cut off levels for EC. Figure 5 shows three cut off values for EC, 0.6, 0.8 and 1.0, which related to goals for ground cover of approximately 70%, 60% and 50% respectively.

The significance of probability maps is their ability to display the risk associated with the rehabilitation goal, and enable decision makers to choose a level of risk that is appropriate and identify areas that may require special attention, as well as in some cases identify areas that will not require any remediation.

6. CONCLUSIONS

This study presents a new approach towards the quantification of uncertainty in soil properties and risk assessment for the purpose of land rehabilitation. The approach presented enables the computationally efficient joint simulation of variables and eliminates the necessity for laborious calculations. These methods include the techniques of MAF that decorrelates variables prior to simulation, and GSGS to quantify the variability of soil properties. The above techniques are shown to facilitate risk analysis and assist decision making in mine site rehabilitation in a case study from a coal mine in Queensland.

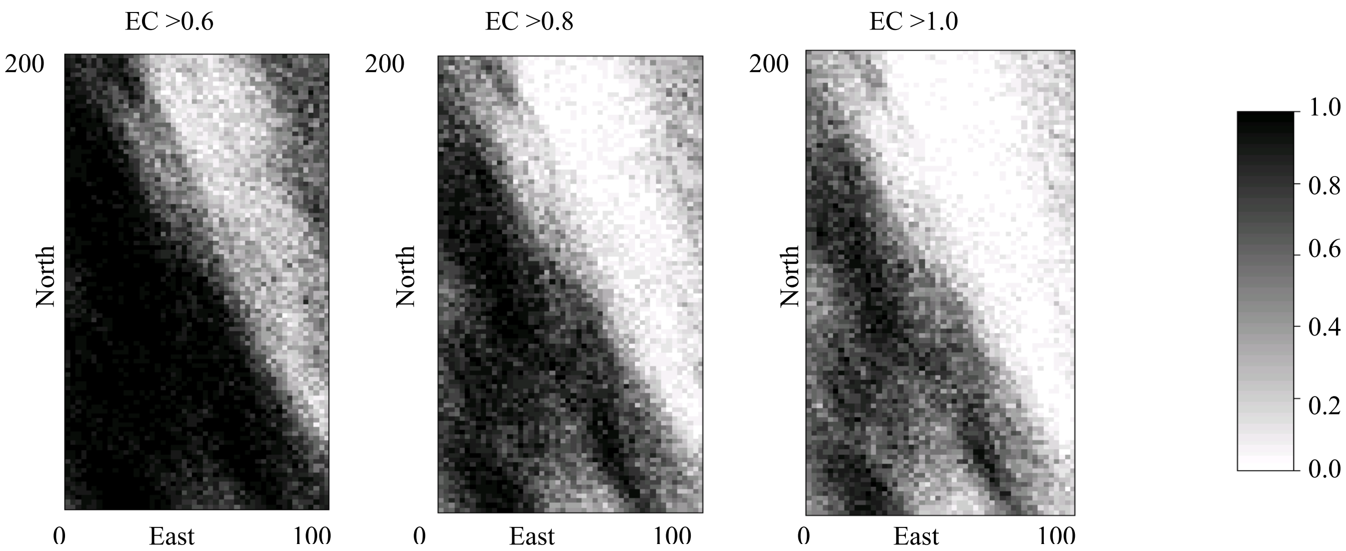


Figure 5. Probability maps for Electrical Conductivity.

7. ACKNOWLEDGEMENTS

This project is an initiative funded by the Queensland Department of State Development through the Sustainable Minerals Institute of the University of Queensland. The measured spoil electrical conductivity and pH data from a mined waste dump was provided by the CMLR and in addition, Dr Andrew Grigg and Dr David Mulligan have provided useful discussions regarding the impact on plant growth of mining waste products in the soil.

8. REFERENCES

Bross, M.J., L. Aarts, C.F. van Tooren, and A. Stein, Quantification of the effects of spatially varying environmental contaminants into a cost model for soil remediation, *Journal of Environmental Management*, 56, 133-145, 1999.

David, M., *Handbook of applied advanced geostatistical ore reserve estimation*, Elsevier, Amsterdam, 216 pp, 1988.

Desbarats, A.J. and R. Dimitrakopoulos, Geostatisical simulation of regionalized pore-size distributions using min/max autocorrelation factors, *Mathematical Geology*, 32(8), 919-942, 2000.

Dimitrakopoulos, R. and X. Luo, Generalized sequential Gaussian simulation on group size ν and screen-effect approximations for large field simulations. *Mathematical Geology, in press,* 2003.

Goovaerts, P., Geostatistical modelling of uncertainty in soil science, *Geoderma*, 103, 3-26, 2001.

Grigg, A., M. Shelton, and B. Mullen, The nature and management of rehabilitation pastures on open-cut coal mines in central Queensland. Tropical Grasslands, 34, 242-250, 2000.

Grigg, A.H., B.R. Emmerton, and N.J. McCallum, The development of draft completion criteria for ungrazed rehabilitation pastures after open-cut coal mining in central Queensland, *Final Report ACARP Project C8038,* Centre for mined land rehabilitation, University of Queensland, 2001.

Gutjahr, A., B. Bullard, and S. Hatch, General joint conditional simulations using a fast Fourier transform method, *Mathematical Geology*, 29(3), 361-389, 1997.

Morrey, D.R., Principles of economic mine closure, reclamation and cost management, *Remediation and management of degraded lands,* Lewis Publishers, Boca Raton, 1999.

Pachepsky, Y., and B. Acock, Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates, *Geoderma*, 85, 213-229, 1998.

Wackernagel, H.J., *Multivariate geostatistics*, Springer, Berlin, 256 pp, 1995.

Webster, R., and M.A. Oliver, Optimal interpolation and isarithmic mapping of soil properties: VI. Disjunctive kriging and mapping the conditional probability, *Journal of Soil Science*, 40, 497-512, 1989.