GEOSTATISTICAL MODELING OF ACID ROCK PREDICTION UNCERTAINTY[1]

Mario E. Rossi[2]

Abstract. The estimation of actual or potential acid rock drainage (ARD) at mine sites is usually accomplished by sampling specific parameters that allow detection and prediction of the potential for ARD. The use of block models to estimate and describe the spatial extent of relevant variables is becoming more common, although quantification of the uncertainty associated with the problem is generally not available, yet it can be critical in an ARD characterization study.

Uncertainties in sampling and analytical processes, in the characterization of the volumes and areas affected or potentially affected by ARD, in the interpolation of sampled values, and in the characterization of physical processes that allow prediction of fate and transport, are always present. It is unrealistic to pretend that the estimation process is error-free, and thus it follows that it is important to provide adequate models of uncertainty, in addition to reasonable estimates of ARD potential. The model of uncertainty can then be used to develop technical risk assessments, including false positives or negatives of certain variables exceeding (or not) certain thresholds.

This paper outlines a stochastic method based on geostatistical conditional simulations that allows assessment and modeling of uncertainty in spatial modeling. This assessment is then translated into risk levels, allowing for a decision-making process that is based on levels of uncertainty. The concept of Loss Functions is illustrated with an example drawn from a porphyry Cu-Mo deposit in South America.

Additional Keywords: Acid Rock Drainage, Conditional Simulations, Uncertainty Model, Loss Functions, False Positives, False Negatives, Risk Models.

[1]Paper presented at the 7th International Conference on Acid Rock Drainage (ICARD), March 26-30, 2006, St. Louis MO. R.I. Barnhisel (ed.) Published by the American Society of Mining and Reclamation (ASMR), 3134 Montavesta Road, Lexington, KY 40502

[2] Mario E. Rossi is the Principal Geostatistician with GeoSystems International, Inc., Delray Beach, FL, 33446.

7th International Conference on Acid Rock Drainage, 2006 pp 1717-1727
DOI: 10.21000/JASMR06021717
https://doi.org/10.21000/JASMR06021717

Introduction and Objectives

Characterization of ARD is now a typical component of any Feasibility Study, remedial investigation, or closure planning in any mining project around the world, and requires a committed and time-consuming effort. This investment is orientated initially towards assessing the potential for ARD in the short and long term. Mitigation and minimization of future negative consequences of contamination on human population and ecosystem are also considered. The results of such mitigating effort are generally measured in terms of eco-health and, sometimes, human health risk assessment. During this process, assessment of the risk associated with estimation errors is almost always lacking.

ARD characterization is usually accomplished by gathering field data of different types, such as neutralization potential, total sulfide, sulfur (or rather, SO_4^{-2} sulfur), carbonate and carbon dioxide, as well as water quality analyses and specific trace elements. Several authors have discussed the use of field samples and blocks models to characterize different aspects of ARD-related problems, see for example Miller and Hertel (1997), Downing and Giroux (1993), and Downing and Madeisky (1997).

In practice, several factors contribute to technical risk, including but not limited to:

- Uncertainties in the initial sample collection, which are related to sampling techniques used, the specific locations that are sampled and not sampled (i.e., sample location biases), and the relationship between sampling methods and the heterogeneity of the elements being sampled. A common example is not sampling all the important rock or mineralogical types that may have an impact in ARD estimation and prediction.
- Uncertainties related to sample preparation and analysis, see for example Gy (1982) and François-Bongarçon (1999), among many others.
- Errors stemming from inadequate handling of data, including data entry and database management processes. These errors can be minimized using appropriate data handling and data quality objectives protocols.
- Limitations related to overly simplified or inappropriate data evaluation and modeling techniques, statistical analysis, etc. These include ignoring or overlooking significant sources of spatial and natural data variability.
- Measurement errors related to the concentrations of the variable being analyzed. In many instances, the acceptable contaminant levels are very close to their laboratory detection limits (MDLs). This may introduce a significant technical challenge at the time of sample analysis, since for most methods the accuracy and precision of the analysis degrades near the MDL. The concept of Practical Quantitation Limit (PQL) has been proposed to overcome this problem, see for example Gibbons (1994).

Given the imprecise information handled, the overall uncertainty in the prediction processes involved may be significant. This uncertainty should be modeled, and should include uncertainties related to sampling and assaying, as well as uncertainty associated with the spatial modeling of the variables.

The objective of this paper is to present an alternative that allows to model uncertainty and quantify the potential risks related to prediction uncertainties.

Method

This paper proposes the use of geostatistical conditional simulations and the concept of Loss Functions to model the uncertainty involved in ARD assessments. These spatial stochastic simulation tools have become in recent years the preferred toolbox for uncertainty modeling and spatial data analysis for mining and petroleum applications.

Many of the statistical techniques sometimes used to analyze environmental data are based on stringent assumptions about statistical distributions, lack of spatial correlation, and independence among the samples considered. These are typical requirements of Gaussian-based statistical techniques often used, for example Analysis of Variance (ANOVA) and Cochran's approximation to the Behrens-Fisher t-test, see among others Gibbons (1994), Gilbert (1987), and USEPA (1989). Therefore, these techniques are inadequate in spatial statistics, where correlation between different sampled points is known to exist.

Geostatistical Conditional Simulations

A general background of the theory of geostatistical conditional simulations is given in Goovaerts (1997) and in Journel (1988). The simulations are models that reproduce the full histogram and spatial continuity of the original conditioning data, and therefore, they honor the spatial characteristics of the variable as represented by the sample data. In addition, it is possible to extend the use of these spatial statistical tools to the time dimension, see for example Rossi and Posa (1991).

By honoring the histogram, the model correctly represents the proportion of high and low values, the mean, the variance, and other spatial statistical characteristics of the data. By honoring the variogram it correctly portrays the spatial complexity of the variables, and the connectivity of low and high contaminant zones. These are fundamental variables that need to be considered in order to improve predictions and diminish predictive uncertainty. When several simulated images are obtained, then it can be said that a model of uncertainty has been obtained.

Conditional simulations are built on fine grids, as fine as possible given the hardware available, so that they correspond to approximately the support size of the original samples. The vertical resolution of the grid is a function of the support data, typically the size of the sampled or screened interval. Larger grid sizes may still be used sometimes because of the amount of computer time and hard disk space involved. In building a conditional simulation model, many of the decisions necessary in typical geostatistical estimations are required, most importantly regarding the definition of the simulation domains (stationarity). Changes in geologic or hydro-geologic domains require splitting the data into different populations. Boundaries between simulation domains can be hard (no data influence across the boundary) or soft, where some data is used from the neighboring domain. Thorough understanding of the behavior of extreme and outlier values in the sampled population is required. Issues such as limiting the maximum simulated grade should be carefully considered.

The simulation method itself should be decided based on the statistical characteristics of the variable being simulated, the quantity and quality of available samples, the availability of using fuzzy information such as geologic descriptions, and the desired output. The most commonly used methods are the Sequential Gaussian (Isaaks, 1990) and Sequential Indicator (Alabert, 1986). The latter is more complicated, is based on multiple indicator kriging techniques

(Journel, 1988), and requires the definition of several indicator cutoffs. The former is simpler and quicker, although more restrictive in its basic assumptions.

All available geologic and hydro-geologic information can and should be used, typically taking the form of "soft" or imprecise information. For example, statements such as "Rock type A is highly acid generating" can be used as prior probabilities in a Bayesian sense, although the details of how to include imprecise information is beyond the scope of this paper.

As with any geostatistical estimation exercise, variogram models should be obtained. These may be particularly problematic, since sometimes there are not enough field samples to obtain such models. This is a potentially serious issue, but there are a number of alternatives that can be resorted to when developing variogram models. Some of these include judiciously applying prior knowledge about the site, data censoring (or what to do with non-detects, sometimes a high proportion of the total sample population), allowed minimum and maximum data values, number of conditioning data to be used, search distances, and assumed directions of anisotropies.

When a number of these conditional simulations have been run and checked, then, for each point defined in the grid, there is a set of possible values for the simulated variable available. These values are interpreted to describe the model of uncertainty for that point, generally arranged as a posterior cumulative conditional probability curve. Preferably, a large number of simulations are needed to describe this curve better. However, due to practical limitations, a much smaller number, perhaps as small as 20-30 simulations, can be used as an initial approximation. When there is significant conditioning information, these simulated values for each cell will not vary much, meaning that the most likely value is known with a good degree of certainty. The opposite occurs when the cell has few samples nearby.

The model of uncertainty obtained for each point can be described as:

$$F(z;x|(n)) = Prob\ \{Z(x) \le z|(n),\ \alpha = 1,...,n\} \qquad \textbf{(1)}$$

$F(z;\underline{x}|(n))$ represents the cumulative conditional distribution frequency curve for each vector $\underline{x}$ of the simulated grid, obtained using the (n), $\alpha=1,...,n$ conditioning filed samples, and it provides the probability of that point in the grid of being above (or below) any contaminant value z.

Loss Functions

Final recommendations in Feasibility Studies and Remedial Investigations (FS/RIs) are typically based on predicted impacts on ecosystems and/or health risk assessments, which in turn are based on estimates of contamination, $z^*(\underline{x})$. Since the true values at each location are not known, errors can and will likely occur. The loss function $L(e)$ (Journel, 1988; Rossi, 1999) is a mathematical function that attaches an economical value (impact or loss) to each possible error, measured in, for example, dollars. If the full set of possible values is known at each location, for example in the form of the conditional probability distribution described in Equation (1), the loss function can be used to obtain the expected conditional loss:

$$E\{L(z^{*} - Z) \mid (n)\} = \int_{-\infty}^{+\infty} L(z^{*} - z) \cdot dF(z; \underline{x} \mid (n)) \quad \textbf{(2)}$$

The minimum expected loss is found by calculating the conditional expected loss for all possible values of the estimates, and retaining the estimate that minimizes the expected loss. As described in Isaaks (1990), the expected conditional loss is commonly a step function whose value depends on the assumed costs of each bad decision, and the relative of costs of miss-classification. This implies that the expected conditional loss depends only on the *classification* of the estimate $z^{*}(\underline{x})$, not on the estimated value itself.

The Loss Function thus quantifies the consequences of false positives and false negatives, weighs the relative impact of each, the probability of each, and then derives the minimum cost solution. For example, in an operation where the mine plan contemplates using the acid neutralizing potential (ANP) of the in-situ rock to influence the scheduling of waste stockpiles, the loss incurred when rock is predicted to be high in ANP when in fact it is not is a direct function of the costs incurred. The cost of the mistakes made can usually be estimated and used to quantify risk. In some extreme cases, when a significant loss of health, quality of life, or life itself results, the cost can be assumed to be infinite. Figure 1 shows a typical Loss Function, where an overestimation error incurs in unnecessary costs, increasing linearly with the magnitude or the error, while an underestimation error causes the Loss to increase exponentially with the absolute value of the error, and for an error of 10.0, it becomes infinite.

An Example from a Large Open Pit Mine

The example described here corresponds to a large open pit porphyry Cu-Au deposit in South America. The purpose of the study was to evaluate the application of the geostatistical assessment of potential risks related to developing waste rock piles with nearly-zero net acid generating potential (AGP). Several variables have been analyzed from drill hole samples, including AGP and acid neutralizing potential (ANP) values, as well as sulfur in sulfates, pyrite, and As. This example is based on ANP, which is defined as the potential for solutes plus particulates in an aqueous system to neutralize acid. It is an estimate of alkalinity, commonly measured in water samples, except that it is taken from non-filtered samples, i.e., includes the acid neutralizing potential of the particulates that may be present. Therefore, it is deemed more representative of the overall acid neutralizing potential of the in-situ rock.

The development of waste dumps with non-acid generating potential may be accomplished by alternatively stacking acid generating rock and rock with high ANP as it comes out of the pit. To accomplish this, a spatial estimate of both variables is required well in advance of mining, such that it can be included in life-of-mine plan and schedule, which in this case is about 15 years, and it implies attempting to estimate in-situ ANP values no less than 200 or 300 meters below current pit surface.

Loss Function, L(e)

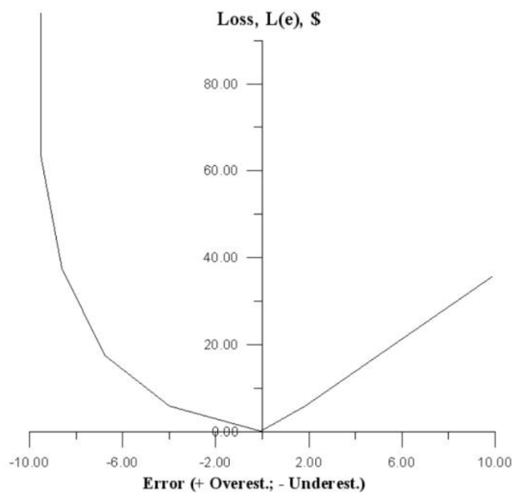

Figure 1: Hypothetical Loss Function, where a positive error is overestimation, a negative error underestimation.

A preliminary analysis of the data available showed significant spatial and temporal variations in water quality as observed in active (pit dewatering) and background (water monitoring) wells, as well as in ex-pit surface water, and highly dependent on seasonal rains. Also, samples taken from exposed surfaces in the pit showed consistent levels of variation. A conditional simulation model was developed for remaining mining reserves (within the designed ultimate pit) for several variables using the Sequential Gaussian simulations (SGS) method. In order to develop the simulations, the following steps were completed:

1. Initial exploratory statistical data analysis was performed over the whole database. This included separating the ANP and other variables by domains, according to their geologic and statistical characteristics. Among those, there are three domains with significant amount of gypsum, in the center and towards the periphery of the deposit.
2. Variogram models were obtained for each variable within each domain. In some cases, due to data scarcity, only an omni-directional variogram was modeled. The models (not shown here) evidence good spatial correlation within some units, with a relative nugget effect between 20 and 40% of total variance.
3. The simulation grid was defined on a 5 x 5 x 5m cell, and 30 simulations were obtained. These simulations provide the model of uncertainty of Equation **(1)**. Figure 2 shows a plan view at level 330m of four of the 30 simulations representing acid-generating

potential[3]. The general spatial trends are reproduced in all simulations, although there are variations in the vicinity of higher values from simulation to simulation.

4. The simulations models were properly validated using the original data and all other simulations parameters chosen.

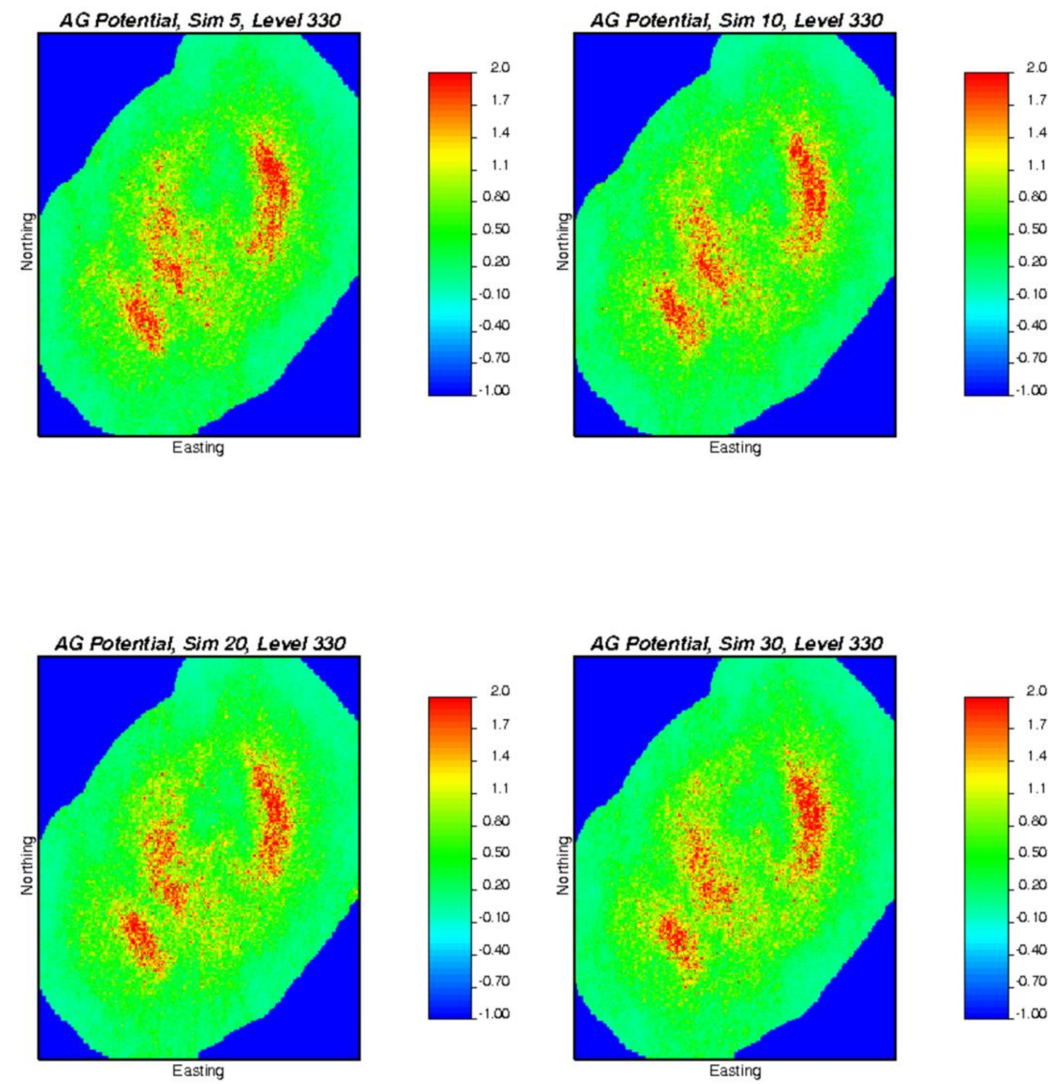

Figure 2: Four AGP simulations, pit level 330m, in meq/l.

To visualize this model of uncertainty, several options are available. One possibility is to use probability maps, such as the one shown in Fig. 3. In this case, the probability that the ANP variable be less than 4 milliequivalents per liter (meq/l) is shown for the same level 330m, which

[3] All values shown in this paper have been factored to protect confidentiality.

in this example is considered a critical value. Note that those areas with high probabilities are almost certainly acid-generating; the interesting areas are those where probabilities are close to 50% (green shades in Fig. 3), where there is little certainty one way or the other. The spatial trends observed are consistent with known geology, including the structural, alteration, and lithology models, including gypsum content. Also, the high probability area to the southwest of the picture corresponds to a simulation domain with few samples, and requires further confirmation.

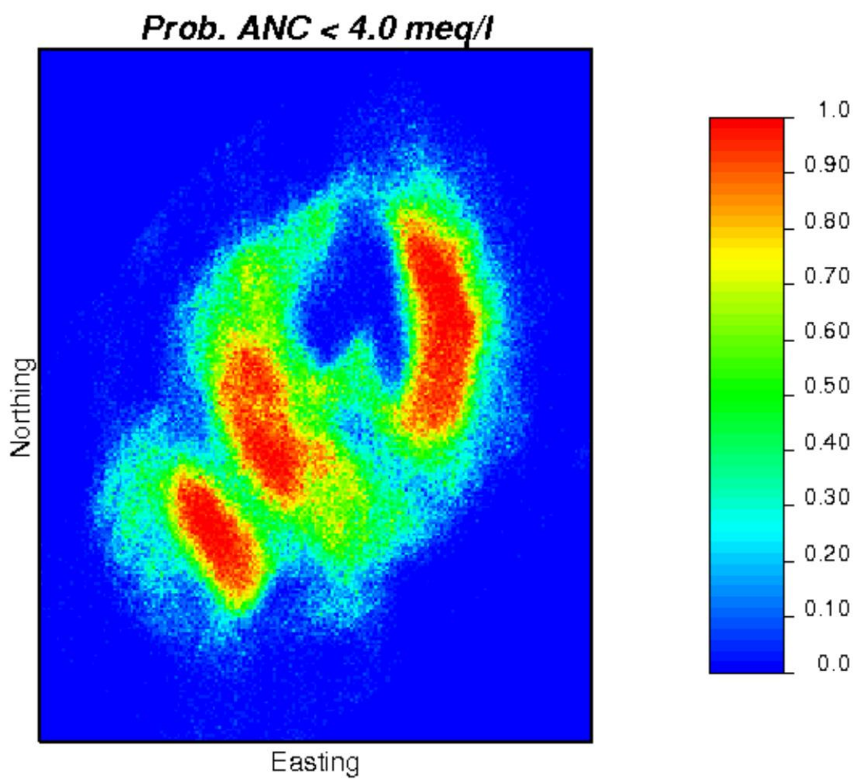

Figure 3: Probability map of acid neutralizing potential less than 4meq/l, pit level 330m, derived from the 30 simulated AGP models.

Developing the Loss Function

The Loss Function applied to evaluate risk in this case was based on the following Equation:

$$Loss = Actual\ Cost - Potential\ Cost \quad \textbf{(3)}$$

The general expression for the costs associated with each type of error are depicted in Fig. 1, where it is assumed that the costs of mishandling acid-generating rock increases exponentially with the error magnitude (which includes an unknown future liability), while the cost of neutralizing rock that did not needed to be neutralized increases linearly with the error magnitude. This amounts to penalizing more underestimation than overestimation, and is a conservative position to take because the operator prefers to be safe and take little or no risks by ensuring that no acid is generated from future dumps.

Results

Applying the loss function described, it is possible to find out the actual economic losses for each simulated value in each cell of the study area, based on probabilities derived from the simulation models. Compositing these losses according to Equation **(2)** result in a mapped "optimal loss" classification. The loss map shown in Fig. 4 is based on the simulations and the specific economic conditions assumed. The loss function suggests that, based on the cost of underestimation and overestimation assumed, it is better to neutralize a larger volume of rock, even if the likelihood of it being acid generating is very small.

The differences between Fig. 3 and 4 depend on the shape of the assumed loss function. The associated risk for each type and level of error is not generally directly proportional to the probability of making the error, except when the loss function is linear for both error types. In this case, a probability measure from the conditional simulations (model of uncertainty) provides a direct measure of risk, which would make Fig. 3 and 4 similar.

Interpretation

This brief example is shows the impact of applying the loss function method described. Traditional estimates provide a "best" estimated value based on nearby samples, but do not provide any measure if uncertainty. In the process of planning to avoid ARD issues, a model that provides not only an estimate but also a measure of uncertainty, i.e., what could the error of the estimate be, can be used to assess the risk resulting from the modeled uncertainty. Estimates such as kriging (in any of its forms) provide an estimate (called a minimum-variance estimate) which implicitly assumes that the consequences of estimation errors are only a function of the absolute error value, and the same whether it is over- or underestimation (Journel, 1988). Srivastava (1987) presents a very lucid discussion on the shortcomings of the minimum variance algorithms.

The situational, political, and socio-economic factors involved in any given case influence the risk tolerance of each mining operation and stockholders to ARD-related risks. The loss function method, while based on a somewhat subjective economic function that requires important assumptions, provides a way to incorporate the degree of risk tolerance specific to each situation.

Maps such as the one presented in Fig. 3 allows determining the volume of material that can be considered acid generating based on the probability of the ANP variable being above a certain threshold, which is an improvement over the use of single estimated value for decision-making. However, applying the loss function adds the possibility of measuring, in hard dollars, the consequences of the predicted error levels at each location, better assessing possible "what if" scenarios.

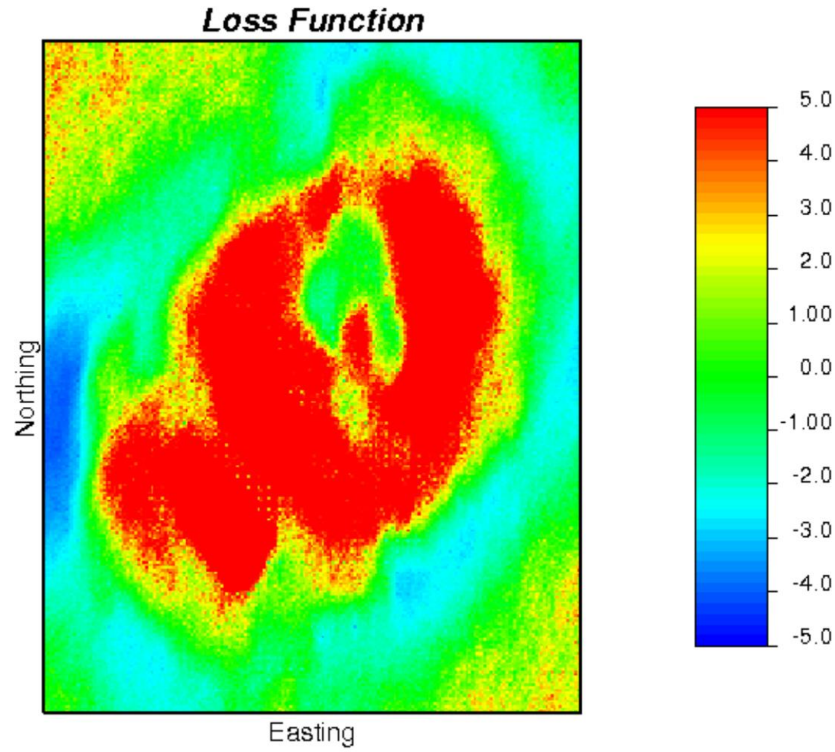

Figure 4: Loss function map (in $ terms), pit level 330m. Red means potentially high loss due to acid generation.

Conclusions

When trying to model ARD-related variables, make decisions, and eventually operate and monitor a prevention or mitigation program, it is often difficult to accurately assess and predict a number of technical aspects of the problem. Most of these difficulties stem from intrinsic spatial and temporal variability of rock quality, sampling inaccuracies, and methodological errors. These variations can lead to mistakes in the decision-making process and may have important consequences. A method has been proposed here whereby the modeled errors are incorporated into the technical risk evaluation process through the use of stochastic conditional simulations, interpreted as models of uncertainty. This requires going beyond the use of block models as has been proposed in the past while attempting to not only estimate the values of relevant variables, but to also provide a model of uncertainty.

These models of uncertainty are then used to evaluate the consequences of all possible mistakes through the use of Loss Functions. Evidently, the quality of the final product will depend on the virtues of the model of uncertainty, and the accurate reflection of incurred additional costs through the Loss Function defined.

A major advantage of this method is its flexibility with respect to assessing costs, since in the formulation of the Loss Function there can be several types of costs included, such as the actual monitoring and mitigation operating costs, costs stemming from health risk assessments, other costs that would be more speculative, including socio-political costs. The cost of such flexibility is a more mathematically involved methodology, and the responsibility that results from actually explicitly stating the hidden assumptions that are inherent to any risk assessment process.

The use of the method outlined here is a risk-based decision-making and planning, performed based on a modeling technique which incorporates key uncertainties associated with ARD prediction, as well as a quantification of the consequences of potential errors.

Acknowledgements

R.A. Miller and J.M. Rendu are gratefully acknowledged for significantly contributing to improve this paper. All opinions, errors and omissions stated herein are solely the author's responsibility.

References

Alabert, F.G., 1987. Stochastic Imaging of Spatial Distributions using Hard and Soft Data. MSc Thesis, Stanford University, Stanford, CA, 1987.

Downing, B.W., and Giroux, G., 1993. Estimation of a waste rock ARD block model for the Windy Craggy massive sulphide deposit, Northwestern British Columbia. *In* Exploration and Mining Geology, Vol. 2., No.3, pp. 203-215.

Downing, B.W., and Madeisky, H.E., 1997. Acid Rock Drainage Study of the Voisey's Bay Ni-Cu-Co Massive Sulphide Deposit, Newfoundland. *Poster* at International Conference Acid Rock Drainage.

Francois-Bongarcon, D. (1999), Extensions to the Demonstration of Gy's, Formula. *In* Exploration and Mining Geology, vol. 7, Nos. 1 and 2, pp 149-155.

Gibbons, R.D., 1994. Statistical Methods for Groundwater Monitoring. J. Wiley & Sons, Inc., New York. http://dx.doi.org/10.1002/9780470172940.

Gilbert, R.O., 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold, New York.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation, Oxford University Press.

Gy, P.M., 1982. Sampling of Particulate Materials: Theory and Practice. 2nd. Ed., Elsevier Amsterdam.

Isaaks, E.H., 1990. The Application of Monte Carlo methods to the Analysis of Spatially Correlated Data. PhD Thesis, Stanford university, Stanford, CA.

Journel, A.G., 1988. Fundamentals of Geostatistics in Five Lessons. Stanford Center for Reservoir Forecasting, Stanford University, Stanford, CA, USA.

Miller, R.A. and Hertel, T.M., 1997. Mine rock characterization and identification of "neutral" rock at the Zortman and Landusky mines, Little Rocky Mountains, Phillips County, north-central Montana. *In* Proceedings of Fourth International Conference on Acid Rock Drainage, Vancouver, B.C., Canada, May 31-June 6, pp 515-532.

Rossi, M.E., and Posa, D., 1991. Geostatistical Modeling of Dissolved Oxygen Distribution in Estuarine Systems. In Environmental Science and Technology, Vol. 25, No 3, pp. 474 481. http://dx.doi.org/10.1021/es00015a015.

Rossi, M.E., 1999. Uncertainty Management through Stochastic Modelling. Proceedings of the 5th International Conference on Water Pollution, Lemnos, Greece, May 24-26.

Srivastava, R.M., 1987. Minimum Variance or Maximum Profitability?, in *CIMM*, Vol. 80, no. 901, 63-68.

U.S. Environmental Protection Agency, 1989. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities. Interim Final Guidance, April.

U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1-A9, available online at http://pubs.water.usgs.gov/twri9A.