Geometallurgical Modeling at Olympic Dam [1]

Jeff B Boisvert[2], Mario E Rossi, Kathy Ehrig[3], Clayton V. Deutsch [2]

[1] Received _____; accepted _____

[2] Centre for Computational Geostatistics, Department of Civil and Environmental Engineering, 3-133 Markin/CNRL Natural Resources Facility, University of Alberta, Edmonton, Alberta, Canada, T6G 2w2.
[3] BHP
[4] GeoSystems International, Inc.

Corresponding Author:
Jeff Boisvert
3-133 Markin/CNRL Natural Resources Facility,
University of Alberta, Edmonton, Alberta, Canada, T6G 2G4
Work: 01-780-492-6157
Fax: 01-780-492-0249
jbb@ualberta.ca

Abstract
Modeling of geometalurgical variables is becoming increasingly important for improved management of mineral resources. Mineral processing circuits are complex and depend on the interaction of a large number of properties of the ore feed. At the Olympic Dam mine in South Australia, plant performance variables of interest include the recovery of Cu and U_3O_8, acid consumption, net recovery, drop weight index and bond mill work index. There are an insufficient number of pilot plant trials (841) to consider direct 3-D spatial modeling for the entire deposit. The more extensively sampled head grades, mineral associations, grain sizes and mineralogy variables are modeled and used to predict plant performance. A two stage linear regression model of the available data is developed and provides a predictive model with correlations to the plant performance variables ranging from 0.65-0.90. There are a total of 204 variables that have sufficient sampling to be considered in this regression model. After developing the relationships between the 204 input variables and the six performance variables, the input variables are simulated with sequential Gaussian simulation and used to generate models of recovery of Cu and U_3O_8, acid consumption, net recovery, drop weight index and bond mill work index. These final models are suitable for mine and plant optimization.

Introduction
Conventional resource estimation is focused on one or a few metals that are of interest in a particular domain because their extraction will be profitable. Increasingly, it is becoming important to understand other characteristics of the ore that affect processing performance and recovery. The detailed spatial distribution of these variables permits a more holistic optimization of the mining operation. This study relates to BHP Billiton's Olympic Dam project in South Australia. Two important topics are addressed with the wealth of measurements taken at Olympic Dam: (1) recovery and other performance variables are related to measured rock properties through a multivariate regression model and (2) geostatistical models of the key rock properties are constructed by simulation and used to predict plant performance within the site.

The main difference between typical geostatistical modeling and modeling of geometallurgical variables is that it is a highly multivariate problem with complex relationships between variables. In this study there are a total of 204 measured variables that may or may not be related to the six plant performance variables of interest. A short review of potential multivariate modeling techniques is warranted. There are three general approaches (1) model all necessary variables accounting for their multivariate relationships (Almeida and Journel 1994, Hong 2010, Scott 1992) (2) reduce the number of variables to a manageable subset by amalgamation or elimination (Babak and Deutsch 2009a,b) (3) transformation of the variables to be independent and model independently (Hotelling 1933; Johnson and Wichern 1998; Filzmoser 1999; Desbarats and Dimitrakopoulos 2000; Bailey and Krzanowski 2012 with many additional references therein). A combination of these approaches is used throughout this work to predict various plant performance variables.

An additional complexity involved in modeling geometallurgical data is the compositional nature of the variables. A full sample analysis of all elements must sum to

100%. Moreover, a full mineralogy and mineralogical association analysis also sum to 100%. Without considering the constant sum aspect of the available data, typical multivariate modeling techniques would violate the sum to unity constraint. A good review of best practice for compositional data modeling techniques is provided by Pawlowsky-Glahn and Buccianti (2011). Typically these techniques involve a variable transformation such that there are no summation constraints on the transformed variables. The back transformation enforces the summation constraints.

Minerals of interest in this mine include copper, uranium, gold, and silver. In addition to payable elements, there are a large number of variables that affect plant performance and recoveries. This is especially true for multi-metal deposits such as Olympic Dam because there are multiple processing streams, each with unique technical characteristics. Cu is recovered by flotation and then smelting; U by acid leaching the Cu tailings; Ag and Au are by-products in the Cu concentrate. In these processes, gangue mineralogy can be as important as the payable metal grades in predicting recoveries. For example, modeling of hardness variables allow throughput prediction; certain gangue minerals impact acid consumption and U recoveries; and the copper-sulfur (Cu:S) ratio is a critical parameter for smelting. Obtaining a spatial three-dimensional model of these variables is a high dimensional multivariate problem as there are 204 variables to consider. Geometallurgy is an emerging discipline in mining that can be used to predict plant performance in situ, prior to extracting and processing the ore.

Mineral recovery and expected plant performance are difficult to predict because they are influenced by a large number of variables such as mineralogy, grade, grain size, plant operation parameters, etc. This multivariate problem is often oversimplified by using constant recovery factors and plant efficiencies based on past experience and empirical rules. Such methods are acceptable during the pre-feasibility stages of mineral exploration; however, when results of pilot plant trials are available, statistical methods can be utilized to better predict recovery and plant performance. In this study, 841 bulk samples from flotation and leach tests are used for the calibration of a predictive model. The performance variables of interest are recovery of Cu and U_3O_8, acid consumption (used in the leaching process), net recovery, drop weight index (DWi) and bond mill work index (BMWi). The end result is a model that can be used to predict recovery and plant performance based on available geometallurgical data. Finally, these variables are modeled over the entire site using collocated sequential Gaussian simulation. These simulations are used to assess the six plant recovery variables exhaustively with uncertainty.

Section 1: Linear regression for plant performance prediction

Over 200 variables are available to develop a regression model (Table 1). Using all 204 variables with the relatively few number of bulk samples available would result in a model that is over fit to the available data. Redundant and unimportant variables are identified and removed from the modeling process, reducing the number of variables to 112. Through a sequence of hierarchical variable amalgamation steps the variables are condensed into 4 major sub-categories. A linear model based on these 4 amalgamated

variables provides a robust predictive model that is used to estimate potential mineral recovery and plant performance. The data in Table 1 are used as input to a regression model to predict six sparsely sampled plant performance variables: recovery of Cu and $U_3O_{8;}$ acid consumption (used in the leaching process); net recovery; DWi and BMWi. The more extensively sampled input variables can be simulated at all locations in the deposit and used to predict local plant performance with the regression model, allowing for an optimization of mining and plant performance.

Table 1: Data available. Note that each grouping of data is compositional in nature.

Data Type	Description
Head Assays	This data contains the % content of the following elements: Co, As, Mo, Ni, Pb, Zn, Zr, Sr, Bi, Cd, Cs, Ga, In, Sb, Se, Te, Th, Tl
Mineralogy (proportions and grain size)	10 identified minerals make up the bulk of the deposit. These include: Brannerite, Coffinite, Uraninite, Pyrite, Chalcopyrite, Bornite, Chalcocite, Other Sulphides, Acid Soluble Gangue and Acid Insoluble Gangue
Association Data	Thin sections have been analyzed and the complete matrix of associations between minerals is available. This describes the contact area between two adjacent minerals within a single grain of crushed material.

Table 2: Description of predictive models generated.

Model	Input variables	Comments
Full Model	-head assays -10 mineralogy -10x11 matrix of associations -specific gravity	This model represents the maximum data available.
Typical Model	- head assays -10 mineralogy -specific gravity	This is the base case model. Field data will most likely contain these variables
Limited Model	-limited head assays (see Figures 1 and 2) -7 mineralogy variables -specific gravity	Only head assays that have many samples in the available database are considered.

Methodology

A linear regression model is used to predict the plant performance variables. One drawback with a linear regression model is that all input variables are required for prediction. The regression model cannot be applied directly if a single input variable is missing from a sample. For this reason, three regression models are generated (Table 2). Each model represents a decreasing number of input parameters. Selection of the appropriate linear regression model would depend on the available information at each location in the mine. If only a small number of variables are missing, data imputation strategies (Enders 2010) could be implemented to replace the missing variables. This is only recommended if a small number of variables are preventing the use of a more detailed model.

The regression models are based on a large set of input variables. The variables are merged into super secondary variables based on the correlations between variables (Babak and Deutsch 2009a,b). This is done because there are too few sample data available to accurately determine regression coefficients for the 204 input variables available. The final model is a linear regression on four super secondary variables. The detailed methodology consists of seven steps. These steps are expanded upon below.

1. Remove unimportant and redundant variables.
2. Quantile to quantile univariate transformation to a Gaussian distribution.
3. Merge the variables (level 1). This step reduces the 112 input variables to 23 merged variables.
4. Merge the variables (level 2). This step reduces the 23 merged variables to 4.
5. Regression on the 4 variables and prediction of the plant performance variables: DWi, BMWi, Cu recovery, U_3O_8 recovery, acid consumption and net recovery
6. Back transform the estimated variables.
7. Determine uncertainty in the model.

Step 1: Remove unimportant and redundant variables

The number of variables must be reduced to prevent over fitting. Variables are removed from the analysis because (1) they have a low correlation to the six output variables or (2) they are highly redundant with one of the other input variables. A variable was considered to have a low correlation if the maximum correlation to any of the output variables was less than 0.13. A variable was considered redundant with another input variable if it had a correlation greater than 0.94. This reduces the number of input variables to 112.

There are a total of 841 samples available for modeling; however, not all samples contain all 112 variables used in the calibration of this model. Due to the nature of a regression model, it is necessary that all 112 variables be present for a sample to be used for calibration. Of the 841 samples, 328 samples were retained for modeling the full regression model.

Step 2: Normal score data

All 118 variables (112 inputs + 6 outputs) are independently transformed to a standard Gaussian distribution. This minimizes the effect of outliers on the regression models and supports a linear combination of the variables under a multivariate Gaussian assumption (Steps 3 and 4). A visual assessment of the bivariate relationships between the input data indicated very few non-linear relationships; therefore, linear regression was deemed appropriate. Complex non-Gaussian bivariate relationships would motivate the use of a more advanced technique such as stepwise conditional transformations (Leuangthong and Deutsch 2003), minimum/maximum autocorrelation factors (Switzer and Green, 1984), alternating conditional expectations (Breiman and Friedman 1985) etc.

Step 3: Merge variables - reduce 112 input variables to 23 merged super secondary variables.

There is a danger of over fitting the available calibration data if a regression model is constructed on all 112 input variables. Therefore, subsets of the input data were amalgamated to construct super secondary merged variables. These merged variables are linear combinations of a subset of variables and significantly reduces problem dimensionality while also reducing the risk of over fitting. The selection of subsets is based on the nature of the measurements and communication with expert geologists that are familiar with the deposit; similar rock measurements are merged (Figures 1 and 2). The merged super secondary variables are generated by assigning weights to each variable:

$$M(v) = \sum_{i=1}^{n} \lambda_i v_i$$

where *n* is the number of variables to be merged based on the weights from a likelihood calculation. These weights are generated by solving the corresponding matrix for each merged variable and for each of the six output variables:

$$\begin{bmatrix} \rho_{1,1} & \rho_{2,1} & \cdots & \rho_{n,1} \\ \rho_{1,2} & \rho_{2,2} & \cdots & \rho_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,n} & \rho_{2,n} & \cdots & \rho_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \rho_{0,1} \\ \rho_{0,2} \\ \vdots \\ \rho_{0,n} \end{bmatrix}$$

The right hand side denotes the correlation between one of the variables of interest and the *n* input variables to be merged, while the left hand side is the correlation between all *n* variables to be merged.

These correlation matrices may be poorly conditioned with few data. Poorly conditioned matrices are the cause of extreme weights (λ_i) and introduce unwarranted noise in the predictions. To prevent this, the correlation matrices are modified to improve stability. This correction is accomplished by decreasing the values of the off diagonal elements of the matrix, similar to Tikhonov Regularization, which increases the value of the smallest eigenvalue for the matrix and increases stability. The minimum eigenvalue for the correlation matrices was set to 0.05.

The merged variables are a linear combination of N(0,1) variables. Thus, the mean of the merged variables will be 0 but the variance is not 1. The merged variables are standardized by the standard deviation determined from the following classical relationship:

$$\sigma^2(M(v)) = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j Cov(v_i, v_j)$$

Thus, the final merged variable becomes:

$$M(v) = \frac{\sum_{i=1}^{n} \lambda_i v_i}{\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j Cov(v_i, v_j)}$$

Step 4: Merge variables - reduce 23 variables to 4 merged variables.
There are two levels of variable amalgamation. The first level groups related variables into 16 merged variables and retains 7 additional variables for a total of 23 variables. Figure 1 shows the variables used in the limited model, while Figure 2 shows the variables used in the typical and full models.

The second level amalgamation combines the variables into 4 super-secondary variables used for regression: retained variables; head assays; mineralogy and; associations. After testing multiple combinations of parameters, the dual level variable amalgamation was found to produce accurate and stable results in cross validation. Considering a single level of regression on all variables tends to over fit the available data.

Step 5: Regression.
The typical and limited models are generated by regression on variables A, B and C while the full model considers variables A through D (Figures 1 and 2). Regression was performed with both linear and quadratic terms but after a cross validation analysis it was found that increasing the number of terms beyond the linear coefficients resulted in little consistent gain and the linear model is carried forward. The final model becomes:

$$Prediction = av_1 + bv_2 + cv_3 + dv_4$$

Figure 1: Variables used in the limited model.

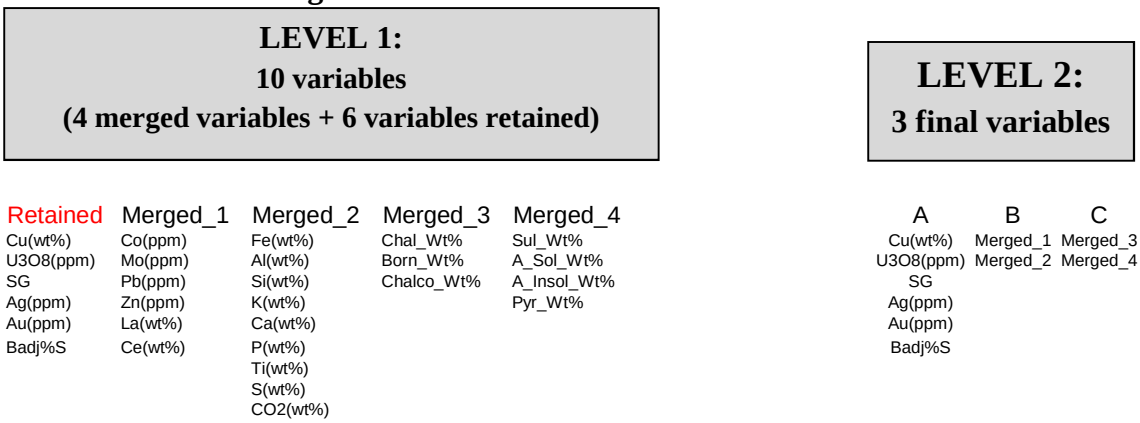


Figure 2: Variables used in the typical and full models.

LEVEL 2: 4 Final Variables

Variable A contains individual variables retained.
Variable B contains the remainder of the head assays.
Variable C contains all mineralogy variables.
Variable D contains all association variables (*not used in the typical model*).

Step 6: Back Transformation.
Once the predictions are made in Gaussian units for each of the six output variables, they must be transformed back into original units using the original transformation tables. The predicted distributions are parameterized by mean and variance values in Gaussian units. Quantiles from these distributions are back transformed; direct back transfrormation of the mean is not possible.

Step 7: Determine uncertainty in the model
The uncertainty in the predicted value can be quantified by modeling the bivariate relationship between the prediction and the truth with a Gaussian kernel (Figure 3). For a given prediction, the potential distribution of true responses can be determined. Consider the difference in making an acid consumption prediction of 60 kg/ton vs. 220 kg/ton (Figure 3). There is clearly more uncertainty in the estimate of 220 kg/ton.

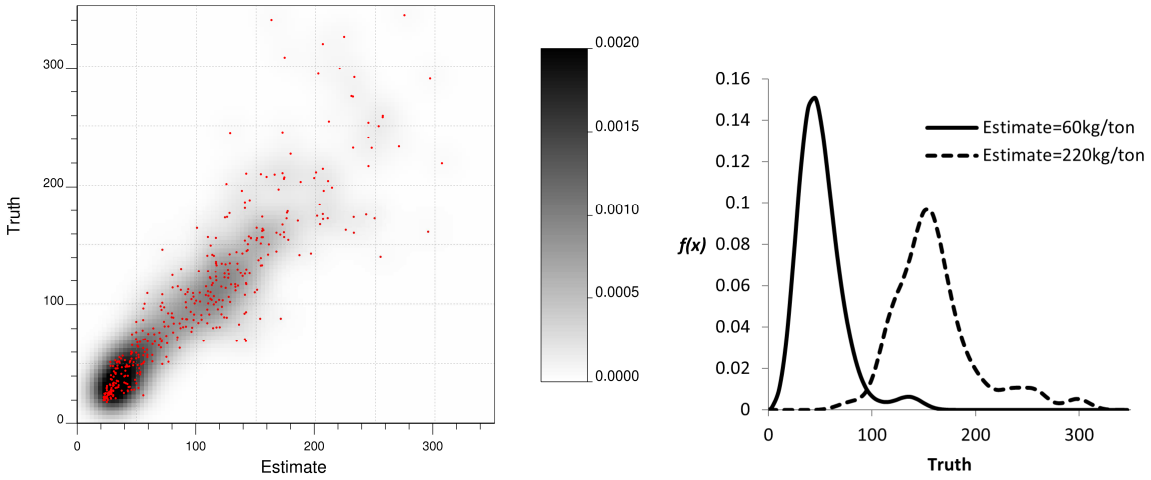
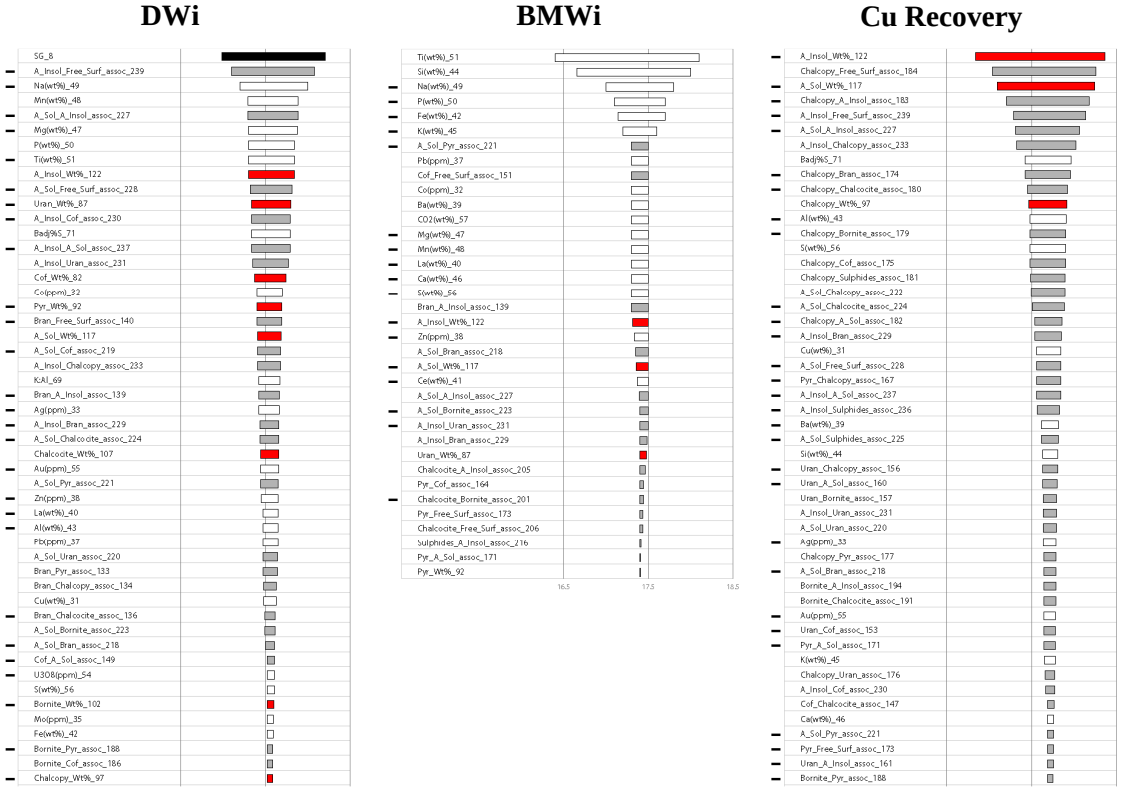


Figure 3: Left: Bivariate density between the truth and estimate for acid consumption. Right: Uncertainty in an estimate of 60 kg/ton vs. 220 kg/ton is shown as the conditional distribution of the modeled bivariate relationship.

Analysis

All samples were used to generate the regression models with the above methodology. High correlation between the estimate and the truth is desirable. Rather than show the 768 coefficients for variable merging and the 24 regression coefficients, a tornado chart (Figure 4) is used to illustrate the influence of each of the 112 variables on the overall model. The lower limit is determined by selecting the p_{10} value for the input variable of interest and setting all remaining 111 variables to their p_{50} value. An estimate is made for each of the six output variables, giving the lower limit on the tornado chart. Similarly, the p_{90} value is selected for the variable of interest to generate the upper limit on the tornado chart. A short horizontal line to the left of the variable indicates that the variable is negatively correlated with the output variable (i.e. the p_{10} response is higher than the p_{90}). Bars are shaded based on the origin of the variable.



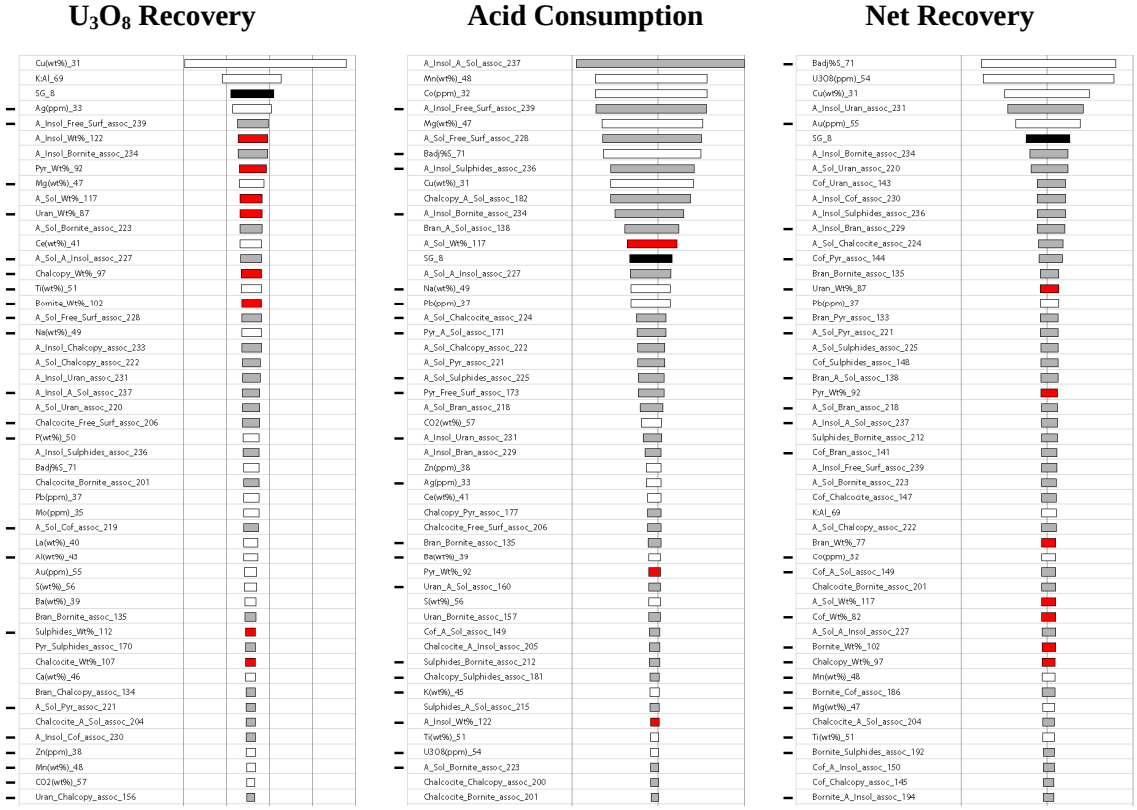


Figure 4: Full model tornado charts for DWi, BMWi and Cu recovery (top) as well as U_3O_8 recovery, acid consumption and net recovery (bottom). White – Head Assays; Gray – Associations; Red – Mineralogy; Black – SG.

Figure 5 shows the predictive ability of the models while Table 3 summarizes the correlation between actual and predicted plant performance based on the regression model. Some interesting relationships were discovered in the cross plots (Figure 5) and the tornado charts (Figure 4):

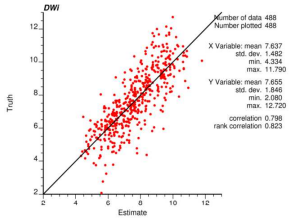
- Na is a significant contributor for DWi/BMWi which indicates different mineralogy.
- SG is important for DWi but not BMWi which is expected because the brittleness of the rock is critical and related to the ratio of iron/silica content.
- BMWi is heavily influenced by the head assays, the top 6 variables contributing to BMWi are from head assays.
- Individual mineralogy variables have little significance (Cu recovery is the exception).
- Presence of Chalcopyrite and acid insoluble gangue are critical to Cu recovery.
- Cu wt% has a large effect on U_3O_8 recovery but little effect on Cu recovery.
- Based on the tornado charts, associations are important for DWI, Cu recovery, acid consumption and net recovery. This is also seen in the comparison of the typical and full models (Figure 5) as BMWi is not significantly altered by removing the association data.
- Recoveries are the most difficult variables to predict (lowest correlation in Table 3). This is expected, as recovery is dependent on a large number of complex interactions.

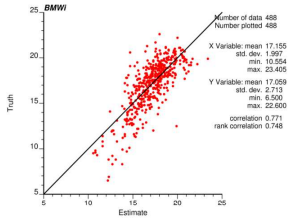
Table 3: Cross Validation correlations for each regression model.

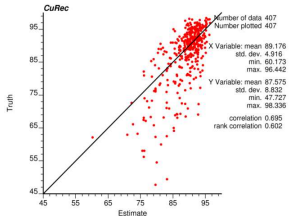
Variable	Limited Model	Typical Model	Full Model
DWi	0.75	0.77	0.80
BMWi	0.53	0.54	0.77
Cu Recovery	0.59	0.59	0.70
U3O8 Recovery	0.61	0.62	0.65
Acid Consumption	0.84	0.86	0.90
Net Recovery	0.67	0.71	0.74

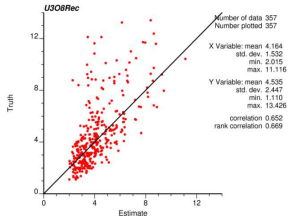
There are a number of opportunities for improvement on the modeling methodology presented in this case study: (1) optimize the merging of the variables at the two different levels. The merging of the variables was done using logical groupings of the 112 variables. An optimization procedure could be developed to select ideal subsets of variables to increase the predictive power of the regression model. (2) Improve the selection of the set of variables to use for each variable predicted. In this work, all 112 variables were used for all 6 output variables. Eliminating some of the less significant variables for individual outputs may reduce noise and increase model accuracy.

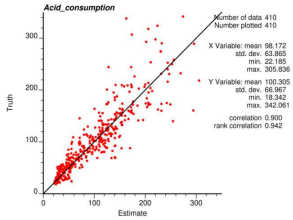
FULL MODEL

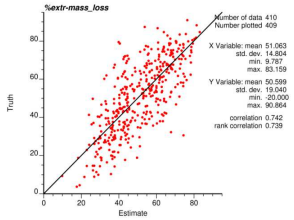




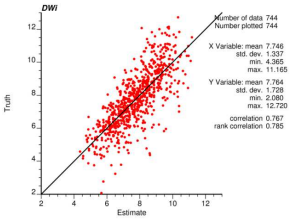


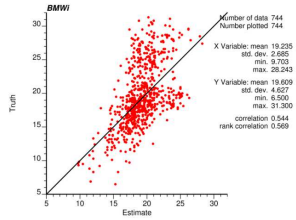


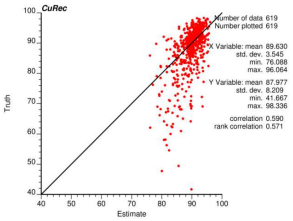


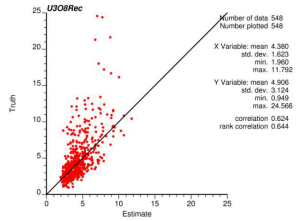


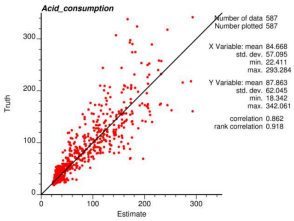
TYPICAL MODEL

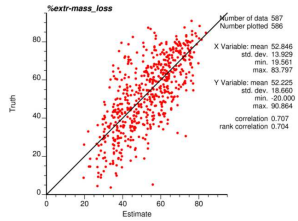












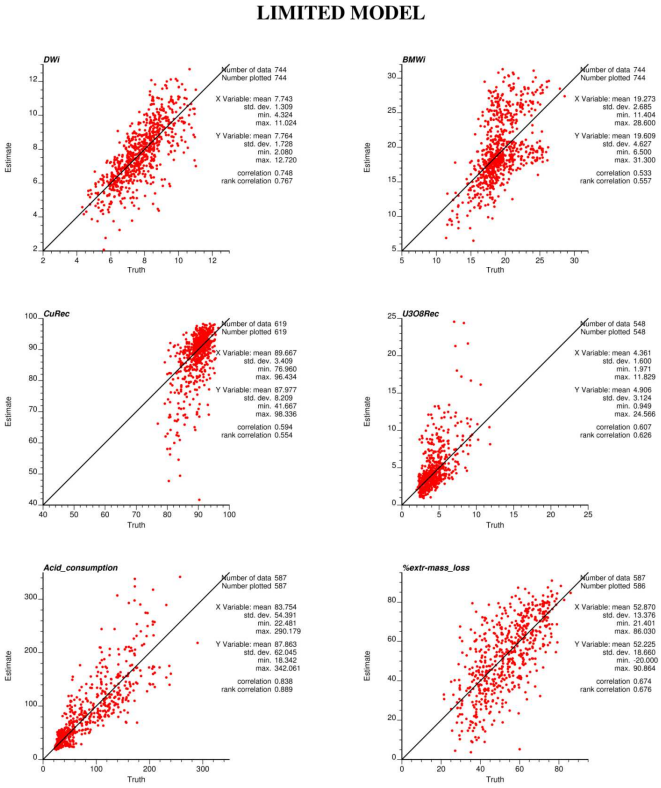


Figure 5: Cross plots of the truth/estimated values based on the full model (above), the typical model (middle) and the limited model (below).

Section 2: Multivariate Compositional Simulation of Non-Additive Geometallurgical Variables

Recovery and plant performance outcomes are influenced by a large number of variables, including head assays, mineralogy and mineral associations. Regression models that utilize all these variables outperform models based on head assays alone. These 112 variables must be spatially mapped in order to utilize the models developed. There are two main difficulties (1) the compositional nature of the variables must be accounted for and (2) many of the variables are correlated and require methodologies that are can be applied to a large number of variables but are effective in reproducing the multivariate relationships. Also note that the model size required to align with current deposit modeling at Olympic Dam requires 27M cells. The computational aspects of generating multiple realizations for 112 variables with 27M cells are significant.

In the proposed methodology, data transformations are used to maintain the compositional nature of the variables and PCA analysis is used to decorrelate the variables. As discussed previously, a multivariate Gaussian assumption is made and PCA fully removes the relationships between variables.

Modeling methodologies are developed for all 112 variables, separated into three groups: head grade assay values; grain size measurements; and mineral associations.

Significantly more samples exist for the head grade variables, therefore they are modeled first. The grain size and association variables are modeled using the head grade realizations as secondary information. This ensures consistency between all variables across the deposit.

The head grade and mineral association data are considered compositional, that is, they are non-negative and sum to 100%. A logarithmic transform is used to deal with this constant sum constraint. Normally, these variables would be co-simulated with sequential Gaussian simulation (SGS); however, the large number of variables and the large model size renders this procedure computationally infeasible. An alternative is to perform a principal component (PCA) transform on the logarithmic data to generate uncorrelated variables. Independent SGS is then preformed on each uncorrelated principal component. The values are back-transformed into original units to generate the realizations. This procedure is used to model the head grade and mineral association data. The grain size data, which are not compositional, are modeled using sequential Gaussian co-simulation for the p_{20}, p_{50} and p_{80} values of each mineral.

Modeling 23 head grade variables

A total of 23 head grade variables are modeled for input into the linear regression models: Cu, U3O8, Ag, Au, Co, Mo, Pb, Zn, Ba, Fe, Al, Si, K, Ca, S, Co2, La, Mg, Mn, Na, P, Ti, Ce. There are a total of 111,572 head assay samples. The K:AL ratio and $B_{adj}S$ are also required, but are calculated from the realizations of K, Al, Ba and S.

The head grade variables are considered compositional because all chemical and mineral rock components must sum to 100%. Not all elements in a sample are assayed; therefore, the sum of the head grades is less than 100%. In geostatistical modeling, if this constraint is not explicitly imposed it can be violated. A logarithmic transform of 24 head grade variables is considered, with the 24^{th} variable imposing the 100% constant sum (23 variables listed above + 1 filler variable). The logarithmic transform is:

$$y_i = \ln\left(\frac{x_i}{x_{filler}}\right)$$

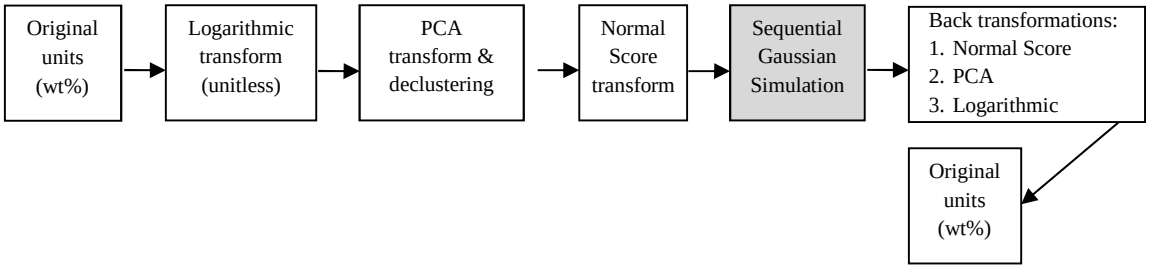
where y_i is the new variable to be modeled and x_i represent each of the original 23 variables. This transformation requires non-zero values for all samples as ln(0) is undefined. The back transformation is:

$$x_i = \frac{e^{y_i}}{\sum_{i=1}^{24} e^{y_i} + 1}$$

There are now 23 logarithmic transformed variables. There are complex relationships among these 23 variables (Figure 6). It would be difficult to reproduce all these relationships with traditional SGS. Therefore, PCA transformation is considered to generate 23 uncorrelated variables. These variables are assumed independent and are

modeled independently with SGS. This ensures good reproduction of the correlation between the 23 variables in the final realizations (Figure 6).

An overall summary of the transformations used is shown below:



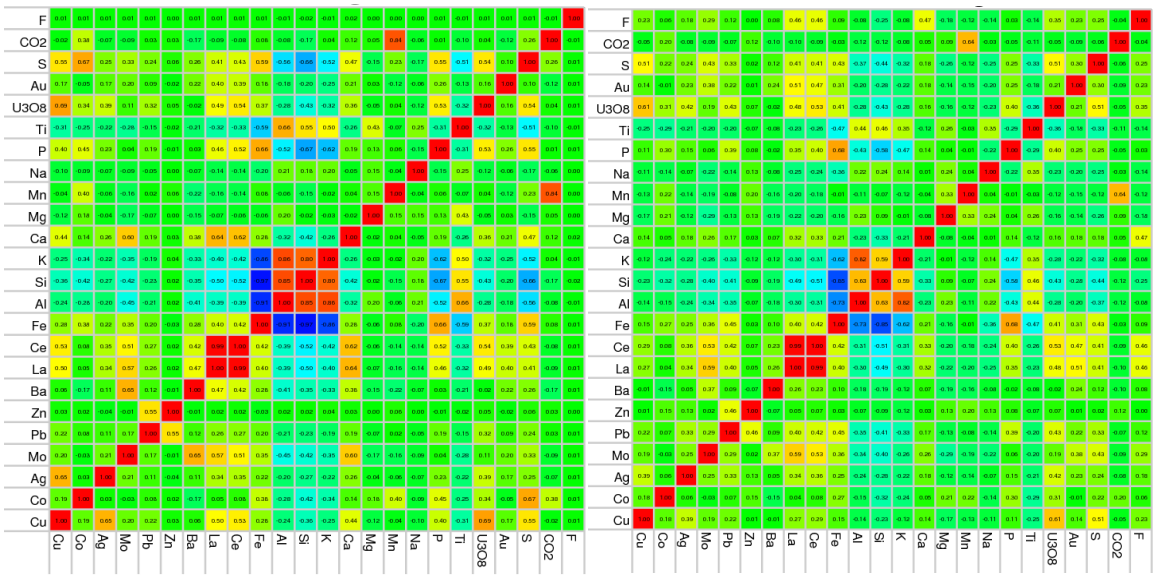


Figure 6: Correlation between the head grade variables (left) and correlation in one simulation (right). Correlations calculated in original units.

Details of Sequential Gaussian Simulation

Implementation of SGS requires the use of variograms for each PCA variable as well as a number of other important parameters. For all variables considered in this case study, simulation was performed with 50 nearby data and parameters for each variogram can be found in Table 4. Because of the large number of variables, variogram fitting software was used with a visual assessment to locate any major inconsistencies.

Declustering was applied to the 23 PCA variables to obtain representative global histograms. A locally varying mean was used in the simulation to consider the non-stationary present throughout the deposit. The local mean for each principal component

was determined with a moving window average of radius 400m in the horizontal direction and 50% anisotropy in the vertical direction.

Table 4: Variograms for the normal score of the PCA head grade variables. A nugget (C0) and two spherical structures (C1 and C2) were used with no plunge angle.

Variable Name	C0	C1	C2	Azimuth 1	Dip 1	Range 1			Azimuth 2	Dip 2	Range 2		
						Major	Minor	Vertical			Major	Minor	Vertical
NS:PCA 1	0.11	0.345	0.544	104	-75	118	79	65	100	-86	1141	1556	548
NS:PCA 2	0.035	0.608	0.357	186	83	67	54	63	158	-56	1417	606	482
NS:PCA 3	0.219	0.348	0.432	360	-80	282	110	197	360	-80	294	1193	945
NS:PCA 4	0.212	0.283	0.505	38	-76	314	79	108	349	-82	530	1627	1488
NS:PCA 5	0.292	0.378	0.33	290	-40	166	166	209	290	-40	670	1449	1303
NS:PCA 6	0.081	0.716	0.202	106	-89	59	54	48	113	-68	535	350	192
NS:PCA 7	0.107	0.302	0.59	50	-76	85	44	55	38	-61	716	1571	947
NS:PCA 8	0.168	0.415	0.417	88	-89	101	60	53	106	-79	471	606	247
NS:PCA 9	0.19	0.455	0.356	89	90	80	64	54	109	-69	496	454	237
NS:PCA 10	0.19	0.545	0.266	311	-12	54	62	73	354	-31	398	210	1020
NS:PCA 11	0.216	0.442	0.342	130	-80	96	68	72	130	-80	550	442	284
NS:PCA 12	0.188	0.426	0.386	281	-16	53	57	81	353	-39	296	247	672
NS:PCA 13	0.239	0.376	0.385	21	83	76	50	55	101	-42	446	713	311
NS:PCA 14	0.201	0.544	0.254	214	-2	49	42	61	224	-45	272	169	290
NS:PCA 15	0.451	0.463	0.085	292	-15	104	141	263	283	24	3791	943	25404
NS:PCA 16	0.234	0.561	0.205	23	-83	68	46	55	44	-58	280	280	784
NS:PCA 17	0.465	0.45	0.085	307	-7	99	122	203	283	-81	43720	1311	35267
NS:PCA 18	0.29	0.424	0.286	198	-5	52	52	67	194	-34	999	374	487
NS:PCA 19	0.211	0.559	0.23	100	-70	55	55	47	145	-73	839	220	148
NS:PCA 20	0.195	0.564	0.241	326	-5	53	57	65	5	-16	684	480	1160
NS:PCA 21	0.332	0.627	0.042	280	-20	51	57	70	280	-20	25464	535	8428
NS:PCA 22	0.305	0.25	0.445	294	-30	81	106	157	281	-61	683	683	365
NS:PCA 23	0.598	0.19	0.212	232	70	142	106	132	231	-53	2037	1022	786

Modeling 9 grain size variables

There are three Uranium minerals of interest: Brannerite, Coffinite and Uraninite. The p_{20}, p_{50} and p_{80} grain size for each mineral has been measured at 497 locations. There is very little correlation between minerals (Figure 7) so each mineral is modeled independently. The correlation between the percentiles of each grain size is reproduced by co-simulating the three percentiles.

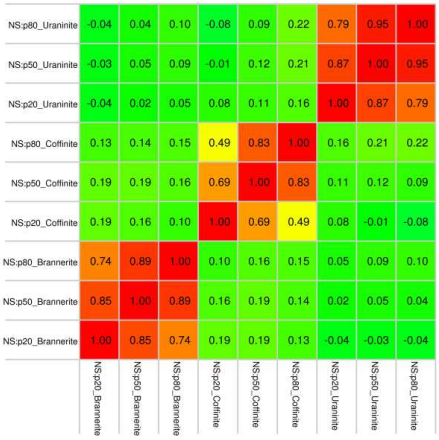


Figure 7: Correlation between the grain size variables. Minerals are simulated independently because of the small correlation between minerals.

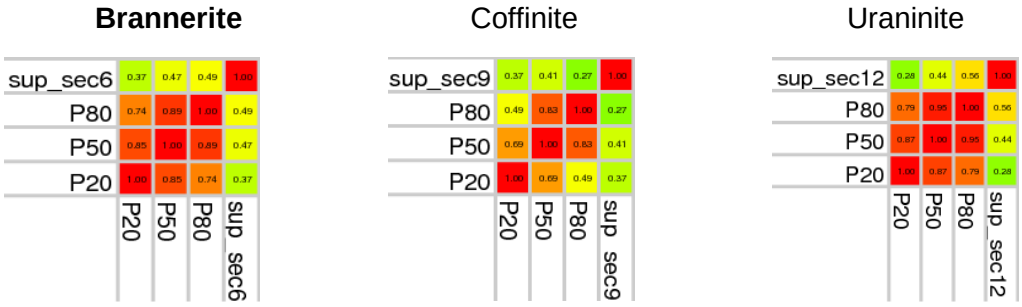
The densely sampled 23 head grade values are used to supplement the lack of information for the grain size variables by considering a super secondary variable which is the amalgamation of the 23 PCA head grade variables. This super secondary variable is created differently for each mineral because the correlations between the mineral grain sizes and the PCA head grade variables differ. To generate this super secondary variable, a linear combination of the PCA head grades is determined from the following equations:

$$\begin{bmatrix} \rho_{1,1} & \rho_{2,1} & \cdots & \rho_{n,1} \\ \rho_{1,2} & \rho_{2,2} & \cdots & \rho_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,n} & \rho_{2,n} & \cdots & \rho_{n,n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} \rho_{0,1} \\ \rho_{0,2} \\ \vdots \\ \rho_{0,n} \end{bmatrix}$$

The right hand side of this equation contains the correlation between one of the grain size variables and the 23 input head grade variables to be merged. The left hand side is the correlation between all 23 PCA head grade variables; note that the left hand side contains 1.0 on the diagonal and 0.0 for all off diagonal terms because the PCA values are uncorrelated. This is done for the p_{50} value for each mineral and the same super secondary variable is used for modeling the p_{20}, p_{50} and p_{80}. This single super secondary variable allows for the cosimulation of the three percentiles and only one exhaustive secondary variable. Without merging all secondary variables into a super secondary, the grain size simulations would have to consider 23 separate secondary variables in order to use all the available information from the head grade variables.

The super secondary variable is used as a collocated secondary variable for each of the grain size models (Figure 8). Note that for the grain size variables neither a logarithmic nor a PCA transformation is considered because there are only three variables (p_{20}, p_{50} and p_{80}) for each mineral. This procedure is repeated for Brannerite, Coffinite and Uraninite. This includes building a new super secondary variable for each mineral.

Very few data exist for the grain size variables and the variograms are unstable. The same variograms are used for the p_{20}, p_{50} and p_{80} of each mineral. The spatial structure for the p_{20}, p_{50} and p_{80} are similar, with the small differences likely due to lack of data. Variograms used are shown in Figure 9 and Table 5.



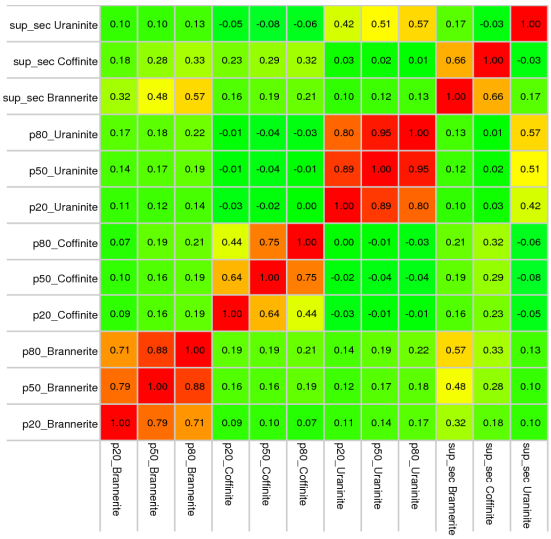


Figure 8: Correlation between the grain size variables. Above – correlations from 497 data to the super secondary variables. Below – correlations from one grain size simulation.

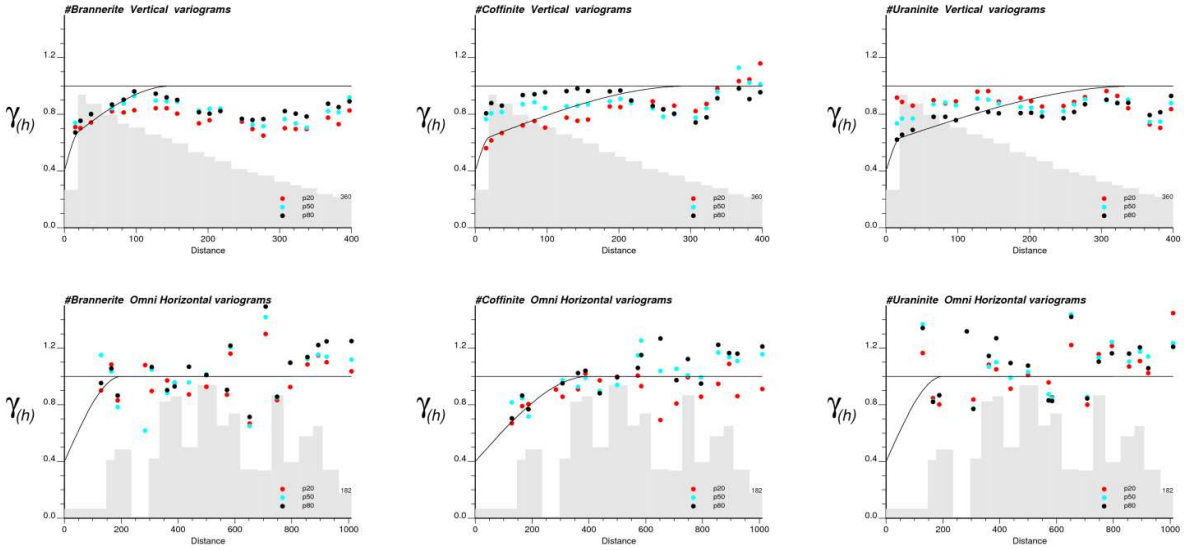


Figure 9: Modeled variograms for the 9 grain size variables. The same variogram was used for the percentiles of each mineral. Background histogram indicates number of pairs for each experimental variogram point.

Table 5: Variograms for the grain size data. A nugget (C0) and two spherical structures (C1 and C2) were used with no plunge/dip angle and no horizontal anisotropy.

Variable Name	C0	C1	C2	Range 1		Range 2	
				Horizontal	Vertical	Horizontal	Vertical
Brannerite	0.4	0.2	0.4	200	20	200	150
Coffinite	0.4	0.2	0.4	400	20	400	300
Uraninite	0.4	0.2	0.4	200	20	200	350

Modeling 100 association matrix variables

Modeling the association matrix utilizes a combination of the techniques previously discussed. The matrix is a 10 x 11 matrix where each row sums to 1.0 (or 100%). Consider this sample:

	Brannerite	Coffinite	Uraninite	Pyrite	Chalcopyrite	Bornite	Chalcocite	Other Sulphides	Acid Soluble Gangue	Acid Insoluble Gangue	Free Surface
Brannerite		8.02								88.18	3.80
Coffinite	1.71		1.64			0.25	0.24		3.50	90.67	2.00
Uraninite		23.51								76.49	
Pyrite											
Chalcopyrite						2.83			2.59	88.43	6.15
Bornite		0.18			0.93				15.50	75.89	7.49
Chalcocite		0.30							0.87	97.91	0.92
Other Sulphides										100.00	
Acid Soluble Gangue		0.05			0.02	0.32	0.01			91.16	8.44
Acid Insoluble Gangue	0.04	0.19	0.01		0.08	0.22	0.16	0.02	12.82		86.45

Each element in the matrix represents the % surface area of interaction between minerals determined from a mineral liberation analysis. Each row sums to 1.0; however, each column does not sum to a constant as the values are standardized by the proportions. There are a total of 100 elements in the matrix, ignoring the diagonals. An assumption that the rows are independent is made to reduce the problem to simulating 10 independent sets (rows) of 10 dependent variables (columns). To maintain the constant sum constraint, the logarithmic transformation is applied to each row resulting in the need to model 9 logarithmic variables. The PCA transformation is applied to reproduce the correlation between variables in each row. The principal components of each row are normal score transformed and then simulated with SGS. There are a total of 490 data available for simulation of the association variables.

As with the grain size variables, the head grade simulations provide a super secondary variable to use in collocated SGS. There are a total of 23 (normal score PCA) head grade simulations to be combined into a single super secondary variable for each of the 100 elements in the association matrix. The PCA transformation is done in such a way that the amount of data explained by each principal component can be measured by the associated eigenvalue. In this case the first 5 components of the head grade realizations contain over 75% of the information in the original head grades. Only the first 5 principal components generated in the head grade modeling are combined into the super secondary variable to reduce the computational requirements of the methodology. Moreover, the super secondary variable is only used for the first 4 of the 9 principal components of the association variables. Because there are 100 association variables to model, available CPU resources are a significant issue.

A variogram is required for each of the 90 principal components (10 sets/rows with 9 principal components in each). As with the head grade variables these variograms were fit with automatic variogram fitting software, visually inspected for inconsistencies and manually corrected when necessary.

Special considerations for the association data

Missing or null values pose a problem in compositional data modeling. In this instance there are entries that are missing because a particular mineral does not appear in a given sample. For rows that have missing values but still sum to 1.0, the missing values are reset to 0.0001 or 0.01%. In some cases there are entire rows that are missing. This is

because the mineral does not appear at that location; however, in these cases all values cannot be set to a small value as they would not sum to 1.0. The solution implemented in this study was to remove the samples where the entire row was missing. When performing SGS at this location the values in that particular row are simulated as if the data did not exist (in fact this data does exist and has a value of zero). The miss-match between the missing values at this location and the simulated values given the surrounding data can be fixed by assigning a 0.0 proportion to the missing minerals, and the mismatched association values accounted for.

Finally, the linear regression models developed in Section 1 are used with these realizations to generate multiple realizations for recovery of Cu and U_3O_8, acid consumption, net recovery, DWi and BMWi which can be used in advanced mine and plant planning/optimization (Figure 10).

Conclusions

Three linear regression models for the prediction of plant performance from head assay, mineralogy and association variables were created. The regression models are used to predict critical plant performance variables from the available samples of head grades, grain sizes, mineralogy and associations. The cost of obtaining samples of plant performance (i.e. pilot plant runs) is very high. Building models based on the sparse sampling of mineral recovery, acid consumption and work indexes allows for the mapping of these variables for all locations in the deposit. This provides a prediction of complex process-based variables that rarely have sufficient data density to generate appropriate variograms and prove difficult to effectively model. The resulting models can used in mine planning and mine optimization including the mine and processing plant together. Limitations of the work include the strong reliance on the multivariate Gaussian distribution after univariate transformation. The sequential approach adopted for the variables and spatially within SGS assumes that the data from previous variables and locations can be successfully transferred through the modeling steps.

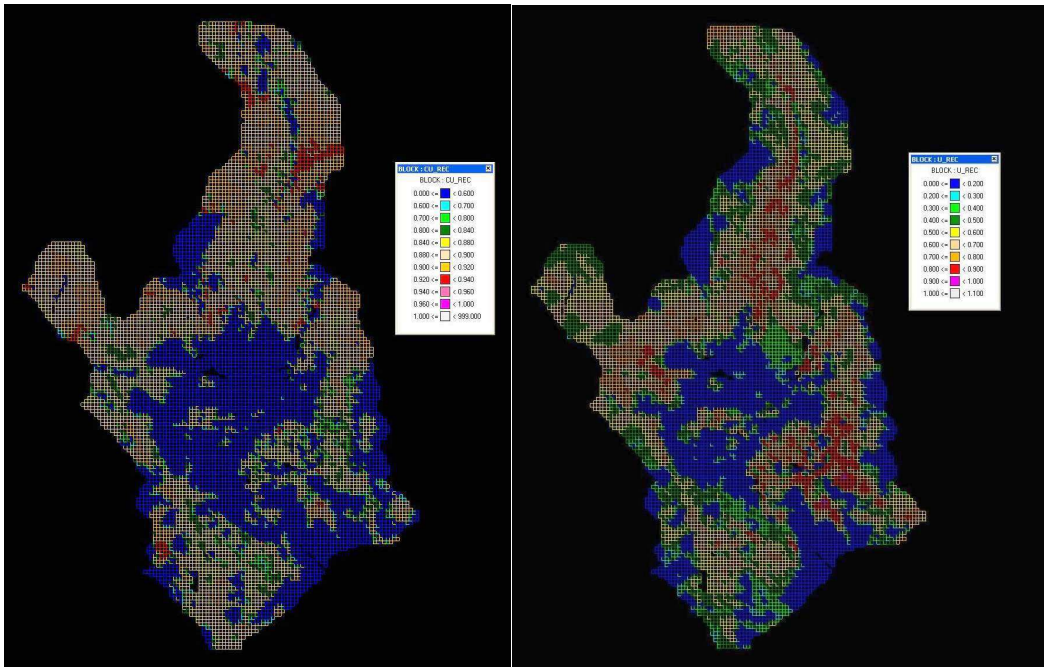


Figure 10: Left: Cu Recovery. Right: Uranium Recovery. -450m Elevation.

References

Almeida, A.S. and Journel, A.G. (1994) Joint simulation of multiple variables with a Markov type coregionalization model. Mathematical Geology, 26:565-588.

Babak, O., and Deutsch, C.V., 2009a Improved Spatial Modeling by Merging Multiple Secondary Data for Intrinsic Collocated Cokriging, Journal of Petroleum Science and Engineering, 69, 93-99.

Babak O. and Deutsch C. V. (2009b). Collocated cokriging based on merged secondary attributes. Mathematical Geosciences. 41:921-926.

Bailey T and Krzanowski W (2012) An overview of approaches to the analysis and modeling of multivariate geostatistical data. Mathematical Geosciences 44:381-393.

Breiman L and Friedman J (1985) Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association. 80(391):580-598.

Desbarats, A.J., Dimitrakopoulos, R., 2000. Geostatistical simulation of regionalized pore size distributions using min/max autocorrelations factors, Mathematical Geology, 32, 919-942.

Enders C. (2010). Applied Missing Data Analysis. New York: Guilford Press. 377p.

Filzmoser, P., 1999. Robust principal component and factor analysis in the geostatistical treatment of environmental data, Environmetrics, 10, 363-375.

Hong S. 2010 …. Multivariate Analysis of Diverse Data for Improved Geostatistical Reservoir Modeling. University of Alberta. PhD dissertation. 188p.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components, Journal of Educational Psychology, 24, 417-441.

Johnson, R.J., Wichern, D.W., 1998. Applied Multivariate Statistical Analysis, 4th edn., Prentice Hall, Jersey, New Jersey, 816pp.

Leuangthong O and Deutsch, C.V. (2003) Stepwise conditional transformation for simulation of multiple variables, Mathematical Geology, 35 (2) 155-173.

Pawlowsky-Glahn V and Buccianti A (2011) Compositional Data Analysis: Theory and Applications. John Wiley & Sons Ltd. United Kingdom. 400p.

Scott, D.W. (1992) Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley and Sons, Inc., New York.

Switzer P and Green A (1984). Min/Max autocorrelation factors for multivariate spatial imaging. Stanford University: Department of Statistics, Technical Report 6, 14p.