Integrating supervised machine learning and geostatistics for enhanced geometallurgical models

J.L. Deutsch,[1] R.M. Barnett[1], and C.V. Deutsch[2]

[1]Resource Modeling Solutions Ltd, Canada
[2]University of Alberta, Canada

Modern geometallurgical data is rich with information spanning physical tests, chemical assays, high resolution visible and hyperspectral logs, mineralogical measurements, and metallurgical test work. While the data available is detailed, it is typically only available on a limited subset of deposit drill holes due to the significant time and expense associated with testing. Supervised machine learning techniques provide powerful tools to integrate geometallurgical measurements into a traditional resource model, generating spatial predictions of the attributes that facilitate mine planning and quantify their associated uncertainty.

There are two primary strategies for integrating geometallurgical data with a resource model using supervised machine learning techniques. Machine learning may be applied on a data scale, learning the relationship of geometallurgical attributes to more widely sampled data such as primary chemical assays and geological logs. These data-scale predictions can be merged with classical geostatistical methods to directly model geometallurgical attributes across the deposit. Alternatively, spatial models of the standard chemical assays and geological units may be generated, and machine learning applied on the block model scale to predict geometallurgical responses.

The scope and scale of both geometallurgical data and traditional geological data available, features of the attributes (compositional constraints, nonlinearity, scale differences), and multivariate relationships all play a role in choosing a strategy for data integration. We propose criteria to consider when evaluating options to integrate supervised machine learning for spatial geometallurgical models and workflows for application.

INTRODUCTION

The goal of resource modelling is to provide predictions of all rock properties of interest at any scale and at all locations within volumes that will be mined. These rock properties include elemental compositions, mineral compositions, mineral associations, size distributions, geomechanical properties including discontinuities, and metallurgical performance variables including recovery, throughput, energy and reagent consumption. Short of providing this ultimate perfection of prediction, we can provide models that capture the intrinsic variability in these rock properties and the associated uncertainty due to relatively widely spaced data.

The available sampling includes more extensively spaced conventional drillholes with the normal assay suite, less extensively sampled geochemical assays and other high resolution scanned variables, and sparsely sampled direct measurements of geomechanical and geometallurgical attributes that are requested by engineers and downstream professionals. These attributes include intact strength properties such as unconfined compressive strength, discontinuity measurements such as rock quality designation and rock mass rating, comminution indices including Bond Work index (BWi) and drop weight index, flotation and beneficiation metallurgical performance measures among others. These properties are collectively referred to as geometallurgical variables or, more simply, *geomet variables*.

There are often one hundred times more conventional assay and logged measurements available than direct geomet measurements. This is reflected in Figure 1, containing an oblique view of drilling data at a synthetic porphyry deposit showing BWi samples focused on the highly altered core of the porphyry with the highest copper grades, and very limited sampling overall compared to the assay data. This example data set will be used to illustrate key workflows in this article. This data was simulated in a hierarchical manner with categorical variables to characterise lithology and alteration. Grades and metallurgical properties were simulated with nonstationary Gaussian techniques to mimic the character of real data that the authors have encountered.

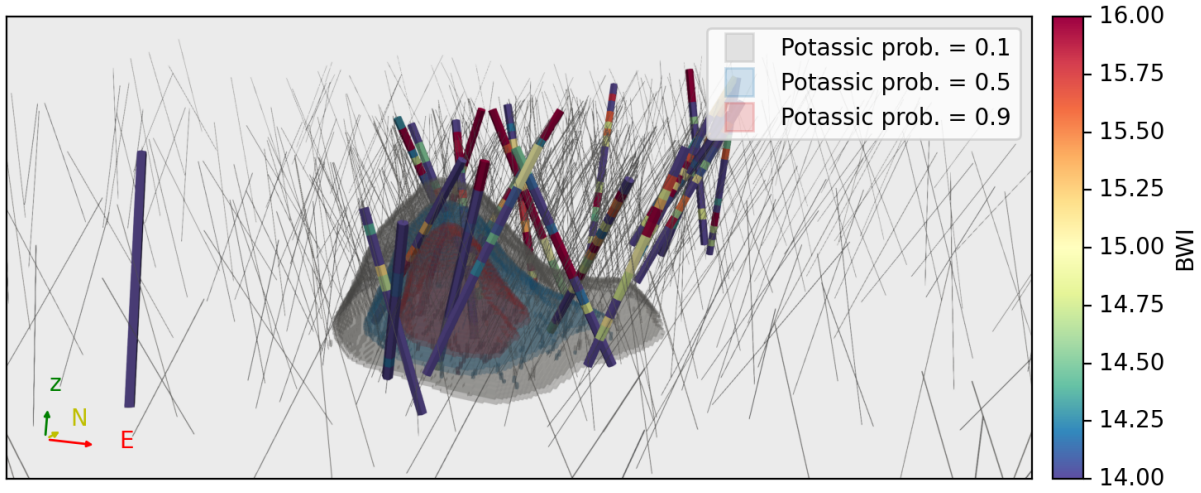

Figure 1. Oblique view of drill hole data at a synthetic copper-molybdenum-gold porphyry deposit. The locations of geometallurgical measurements are coloured by BWi, compared to the locations of assay data in grey.

The challenge of constructing geomet models with limited geomet data is exacerbated by the focus on sampling economic areas driven by the high time and financial cost of conducting geomet tests. As seen in Figure 1, iso-contours of Potassic probability emphasise concentration of BWi sampling in high grade Potassic and the immediately surrounding Sericitic material. Little to no BWi samples are available in lower grade Propylitic material, nearer to the edges of the view. All locations with assay (grey) samples in this figure are relevant to BWi prediction, but sampling is biased towards to the higher-grade areas.

In this article, key unit operations for spatial geomet modelling are described, followed by five workflows for building spatial geomet models, and a discussion of workflow selection. Select results are shown throughout for illustration using the synthetic copper-molybdenum-gold porphyry deposit. Interested readers are referred to Scott *et al*., (2023) for a more detailed case study that implements some of the discussed operations.

UNIT OPERATIONS IN SPATIAL GEOMET MODELLING

All workflows for generating spatial geomet models have similar prerequisites. Stationarity must be established. Exploratory data analysis including univariate and multivariate analysis is performed on stationary subsets of the data to determine appropriate modelling methods. Transformations may be

undertaken including variable reduction or standardisation via a compositional or alternative set of transformations. Further unit operations for spatial geomet modelling will vary from geostatistical methods including kriging, simulation, cosimulation, and Bayesian updating, to ML techniques including tree-based regression and artificial neural networks.

The Decision of Stationarity

Common to all approaches for spatial geomet modelling is the process of determining stationary subsets and performing exploratory data analysis (EDA). The decision of stationarity is primarily the decision of grouping data and the unsampled volume into rational subsets for analysis and modelling, including both geological and geometallurgical properties. The aim is to have stationary domains that are relatively homogeneous within the domains and different between the domains. Geologic criteria such as mineralisation, alteration, lithology, and structure together with the grade or geomet variables being predicted form the basis for the domains. A second aspect of stationarity is the location dependence of statistics. A common assumption is that the joint unconditional probability distribution of geological and geometallurgical values is invariant within the domains. Modern geostatistical and ML algorithms may consider a locally varying mean, standard deviation, and orientation of spatial anisotropy within a chosen stationary domain, so are considered stationary after trend removal.

EDA may be used for identifying major shifts in the geomet responses as a function of combined categorical factors, allowing for geomet domains to be developed. Note that these geomet domains may differ from underlying predictor domains, such as grade domains. Regression models or geostatistical models of the geomet attributes may then be constructed, where the predictor-response relations would have increased stationarity and confidence across each respective domain. The predictors in these regression models would no longer include geological factors that drive domaining, but variability of the geomet responses across domains will reflect their influence.

Exploratory Data Analysis

EDA is a prerequisite for spatial geomet modelling, beginning with standard univariate analysis of attributes and declustering to estimate representative distributions. Multivariate statistics including both correlation matrices and scatter matrices between attributes are required. Spearman (rank) correlations may be more useful in the presence of nonlinearity and outliers than Pearson correlations. These exploratory steps involve frequent discussions across geology, metallurgy, and geomet teams on expected relationships and observed correlations and distributions. These EDA steps are iterative, where these steps will be reassessed during the geostatistical and ML steps of the workflows. Figure 2 displays scatter between the multivariate attributes of the porphyry data. Observe that rank correlation is higher than Pearson correlation for bivariate relations exhibiting non-linear dependence, such as between BWi and specific gravity (SG). This is useful when evaluating many potential predictors and considering reduction schemes as described in the next section.

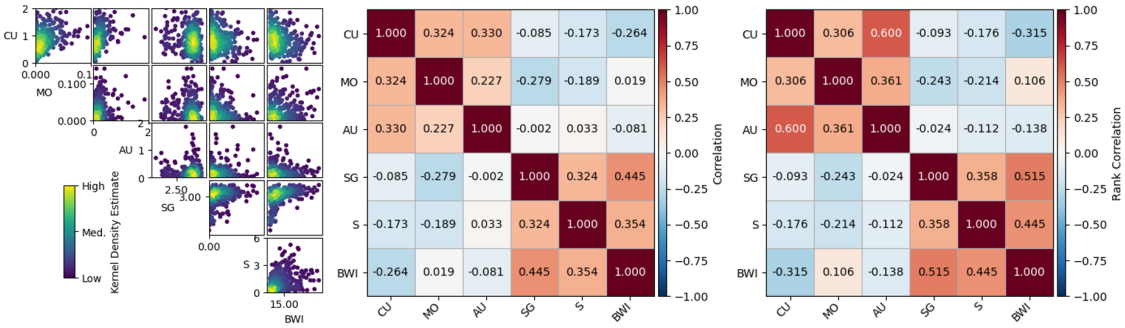

Figure 2. Scatter (left), Pearson correlation (middle) and rank correlation (right) matrices.

Variable Selection and Reduction

Geometallurgical investigations incorporate logged geological factors along with a variety of analyses including scans, geochemistry, and geotechnical information. These variables typically contain

significant redundancy which will be identified during EDA. Geochemical indicators may strongly correlate with alteration or lithology. Geotechnical attributes may correspond with proximity to critical structures. A key component of many geometallurgical workflows is reducing the number of attributes used for modelling to mitigate redundancy and overfitting concerns. This feature selection is typically an iterative process where variables may be removed one at a time with the impact on the ML models assessed via reduction in the coefficient of determination or mean squared error on a k-fold validation study.

Geological factors are often essential predictors of geomet responses and controls on stationarity, including mineralogy, alteration, lithology, and structure. These factors may be considered in one of two applications: (i) use in stationarity related subsetting as previously described, or (ii) direct use in the regression model for predicting the geomet response. Although the geomet regression models would ideally be constructed with stationary response-predictor relationships, this must be weighed against reducing the already low number of geomet data that is available for training the regression models, to even lower numbers within each subset population. Modern ML algorithms provide powerful capabilities with respect to developing accurate predictor-response relationships, while avoiding overfitting in the presence of relatively few records and many predictors. Strategies to mitigate overfitting will vary depending on the number of records and attributes. Nevertheless, subsetting prior to regression modelling will frequently contribute to overfitting concerns and overall degradation of the prediction. This motivates application (ii), where the geological factors are instead used directly in the regression models so that the sparse geomet data is not reduced through subsetting, whilst permitting non-stationarity associated with the input geological factors to be learned and reflected in the response. For example, alteration as a categorical predictor, or deriving simple variables such as 'distance to footwall' that correspond with observed trends, so that associated non-stationarity is correctly reflected in the measured geomet response (comminution attribute, recovery, and so on).

Geostatistical Estimation and Simulation

Geostatistical modelling of geomet variables directly into the spatial model is attractive to ensure realistic spatial variability and fidelity with all available measurements. Linear or additive variables can be averaged, estimated, and predicted with conventional geostatistical tools such as variograms and kriging. Unfortunately, geomet variables often are nonlinear and nonadditive, where the application of kriging or similar estimation techniques would lead to biased (potentially severely biased) estimates.

Simulation or maintaining a distribution of uncertainty are therefore required to avoid averaging. The data transformations and imputation described in the next section are frequently necessary for simulation to reproduce compositional, multivariate, spatial, and non-stationary features of the variables. Post-processing via block-by-block averaging (E-type) or localisation may be required to make the models useable for engineers and metallurgists.

An E-type model represents the expected outcome of each block for the realisation set, thereby providing a conditionally unbiased best estimate that may be useful in certain circumstances where one model is required (Deutsch and Journel, 1998). An E-type model involves averaging, but specialised methods could be used at this late stage of the workflow. An E-type model will not preserve variability of the attributes, which could motivate localisation (Boisvert and Deutsch, 2012) to yield a single model with a distribution that is representative of the realisations, but organising relative highs and low according to the local rank ordering of the E-type model.

Data Transformation and Imputation

Geostatistical and ML methods have strong constraints on the univariate and multivariate characteristics of the data to be modelled. Geostatistical simulation will require univariate normal data. Multivariate simulation techniques may require multivariate normal data with no compositional constraints. Most ML techniques will expect data to be relatively equally distributed across variables to avoid adverse effects on model training.

Prior to applying geostatistical or ML techniques, compositional constraints may be removed by

applying a compositional transform, such as log-ratios or simply ratios (Aitchinson, 1982). The sum constraint imposed on geochemical attributes renders one variable redundant, and the compositional transformation prior to modelling enforces reproduction of that constraint and improves the resulting models (Pawlowsky-Glahn and Olea, 2004). This transformation step precedes the transformations described below.

Additional data transformations are used for transforming the data to be multivariate normal and stationary. The most common workflow that has emerged involves the following steps:

1. Normal score (NS) transform each variable (Deutsch and Journel, 1998), transforming the data to be marginally normal, which generally improves the trend modelling and Gaussian mixture model (GMM) results that follow.
2. Model the trend of each NS variable to capture remnant non-stationarity, commonly using a moving window average or block ordinary kriging. The relationship between each trend and its NS score variable is then parameterised with a GMM, before using a stepwise conditional transform to remove any non-stationary relation (Qu and Deutsch, 2018).
3. Apply the projection pursuit multivariate transformation (PPMT) to remove multivariate dependence that remains between the detrended variables (Barnett *et al*., 2014).

After independently simulating the transformed variables, the transforms are inverted to re-introduce multivariate relations, non-stationarity, and original units to the simulated realisations. A recent example of this workflow is seen at Olympic Dam in Minniakhmetov and Clarke (2023). Note that multivariate transformations such as the PPMT require isotopic (equally sampled) data. Multiple imputation may be used for addressing heterotopic geological data (Barnett and Deutsch, 2015; Silva and Deutsch, 2016), generating realistic isotopic data realisations appropriate for the described transformation workflow, while capturing uncertainty associated with the imputation process.

Supervised Machine Learning

Unsupervised ML searches for patterns and relationships in the data, independent of any specific variable of interest. Such techniques are potentially interesting in some circumstances; however, our goal is to forecast geomet variables and not simply to explore the data. Supervised ML, where a target (geomet) response variable is predicted from a set of more widely available data, is of direct interest to the applications being addressed in this paper.

The more prevalent geochemistry, logging, and geotechnical data available to predict geomet attributes spatially, and at collocated samples, is typically tabular. Considering the example of BWi, SG measurements are typically collected along with the assays at the sampled intervals. This makes the data amenable to many traditional ML techniques, including linear regression, as well as tree-based techniques including decision trees, random forest, and stochastic gradient boosting or multilayer perceptrons (neural networks). The predicted line (a single geomet variable is regressed against a single variable), for a linear regression and regression tree regressor are shown in Figure 3. In addition to better predicting this nonlinear relationship, the tree-based regression limits extrapolation at very high or low specific gravities.

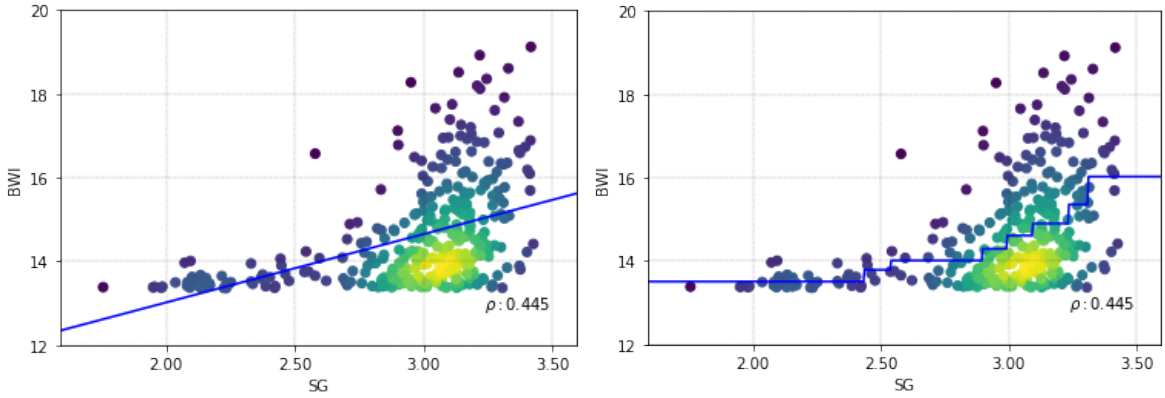

Figure 3. A simple linear regression model of BWi vs SG, and corresponding decision tree regression, right.

Tree-based algorithms, including random forest (RF) and stochastic gradient boosting (SGB), or more commonly the SGB variant extreme gradient boosting underpin many geometallurgical predictions. Regression using a collection of decision trees limits extrapolation, visible in the flat regions of Figure 3 where the decision tree levels out below a specific gravity of 2.4 or above a specific gravity of 3.3. The internal resampling conducted in RF and SGB makes the techniques relatively robust to overfitting with minimal hyperparameters for tuning, making these a practical option for machine learning integration with geomet workflows. The interested reader is referred to Chen and Guestrin (2016) and Friedman (2002) for a comprehensive description of the techniques and details on hyperparameters.

Artificial neural networks (ANN), typically multilayer perceptrons may similarly be applied and are effective for data with a more complex structure, such as images or videos, that are not tabular. Reviewing the implementation is outside the scope of this article, and the interested reader is referred to Kingma and Ba (2014) for a common formulation of stochastic optimising solvers.

Considerations when Training Machine Learning Models

A characteristic of most geomet data sets is a high degree of missing data. As many techniques are destructive and require large volumes of material, only a subset of measurements may be available on a given interval of core. Most ML techniques will require input measurements to be isotopic, with no missing data. Depending on the missing data mechanism (Enders, 2010), imputation of the missing data is a practical approach. Within the machine learning setup, iterative imputation using a simple technique (linear regression, RF) can be pre-applied to predict first the missing values, then those predicted values used in the ML process.

Many geomet data configurations have as many attributes measured as geomet measurements available. This makes the ML process inherently unstable and susceptible to overfitting. The degree of overfitting can be monitored with K-fold validation (Figure 4), where the data is divided into *K* subsets (five is common), where *1/K* of the data is held back as a testing data set. The ML model is trained and predictions compared to evaluate how well the ML model is able to predict on the training (or validation, if available) subset. The authors have found that the double descent (OpenAI, 2019) phenomenon, where ML models with 1000s of parameters fit on only 100s of data exhibit better performance after crossing the interpolation threshold. Figure 5 displays BWi prediction that results from this process for the porphyry example, comparing the predicted and true BWi when training the model on all data (left) or five-fold data (right).

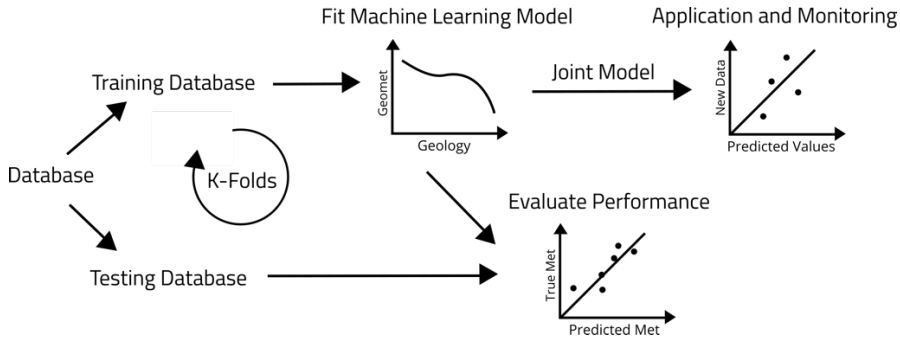

Figure 4. K-fold testing process for iterative tuning of machine learning methodology and hyperparameters.

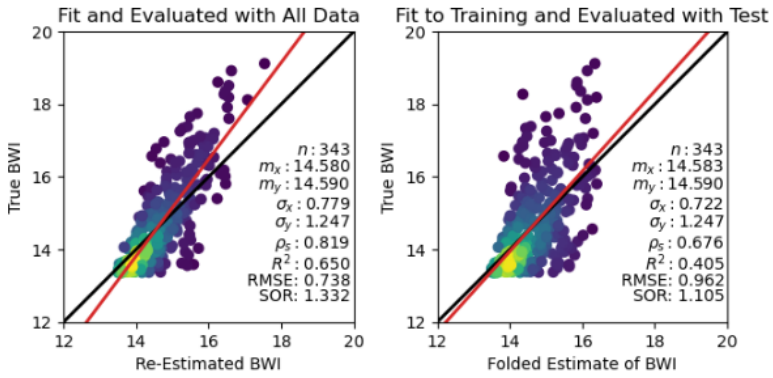

Figure 5. K-fold testing process for iterative tuning of machine learning methodology and hyperparameters.

Explanatory methods

The sparse data and complex relationships require input from geologists and metallurgists to understand model behaviour and determine reasonableness. ML algorithms should be assessed using explanatory techniques such as individual conditional expectation (ICE) plots or partial dependence plots to better understand the impact of individual characteristics on the geomet variables. These are imperfect summaries as they rely on the median or resampled behaviour to decompose a multivariate model to a bivariate explanation, so should not be treated as the ultimate source of model truth, but provide useful information. ICE plots and categorical partial dependence plots are shown in Figure 6, associated with the porphyry BWi prediction from Figure 5.

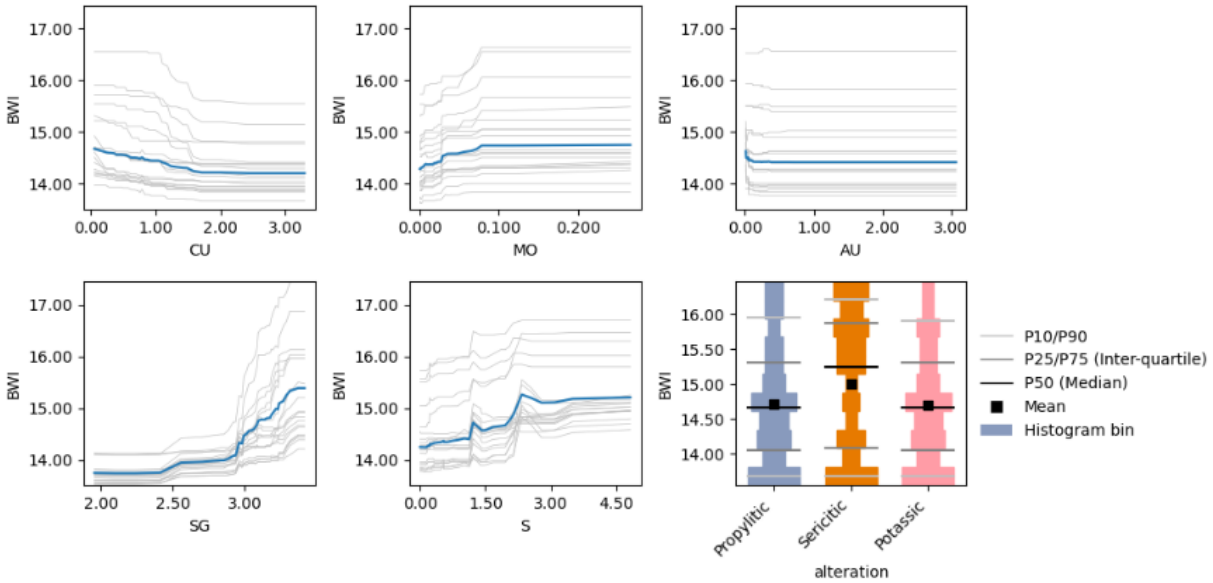

Figure 6. Explanatory model *ICE plots and categorical partial dependence plot (bottom right).*

WORKFLOWS FOR SPATIAL GEOMETALLURGICAL PREDICTION

The described unit operations can be assembled in a variety of techniques to generate geometallurgical models. Five workflows our team has applied are described here along with their key limitations.

Workflow 1: Conventional Machine Learning and Geostatistical Estimation

With the sparse sampling ratio of geomet variables to other measurements, the traditional way of predicting geomet variables based on the more widely available measurements is with regression. That is, an equation or complex relationship is used to predict the geomet response variables as a function of the available predictor variables, where modern ML algorithms are commonly used for determining

the predictor-response relations. Applying random forest, stochastic gradient boosting, or ANN permits the rapid integration of dozens or hundreds of supporting variables to predict the geomet attribute of interest. In the presence of particularly sparse data, a single average value may be assigned to the entire domain.

In this conventional approach, the ML model is applied to estimated geological values in the blocks that are prepared during the resource estimation phase. This approach (Figure 7) is reasonable, however, the spatial continuity of the geomet variables is important for blending and subsequent design, and estimated geological inputs may be too smooth for the ML model. Given the nonlinear relationships typically observed between geological and geometallurgical attributes, this may also introduce a bias in the final geomet block models.

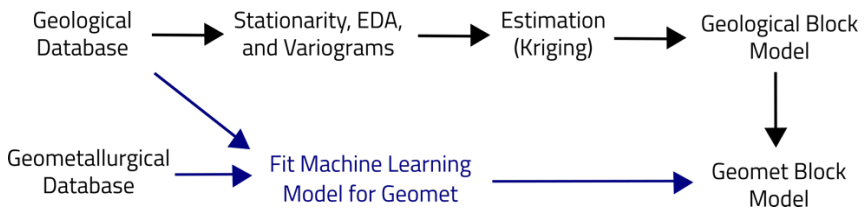

Figure 7. Conventional prediction of geomet variables with machine learning on an estimated block model.

Workflow 2: Direct Simulation Prediction

As an alternative, the geological and geomet variables may be simultaneously modelled using multivariate geostatistical simulation approaches (Figure 8). This approach treats geomet variables the same as grades, density, and geotechnical variables, incorporating them into the same workflow. Given the large scale of many geomet variables relative to the small assay scale, this can require upscaling the geological attributes to a very large composite scale or downscaling the geomet attributes.

This avoids concerns that surround averaging of geometallurgical attributes, leading to loss of variability and potential bias. It does, however, place an extreme dependence on the imputation step, which is relied on to populate the very sparsely sampled geometallurgical attributes across the database. Although the imputation methods may reproduce complex multivariate and spatial relations, they could underperform more complex ML models in terms of accurately predicting geometallurgical attributes. Further, the extreme heterotopic pattern that would result, could compromise convergence of Gibbs-sampler based algorithms used within modern imputation methods for geological variables (Silva and Deutsch, 2016).

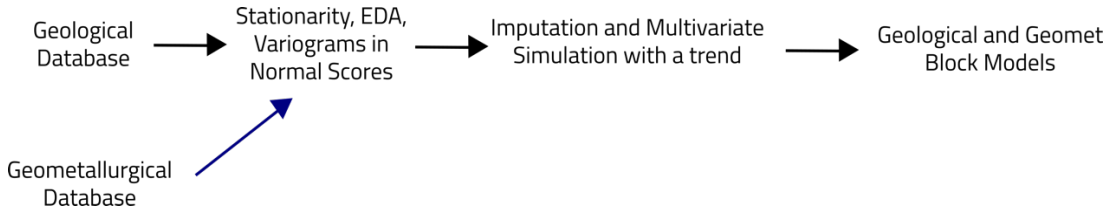

Figure 8. Direct multivariate geostatistical simulation to simultaneously generate geological and geomet models (realisations), which may be post-processed to a single model.

Workflow 3: Post Simulation Prediction

Concerns with the prior two workflows motivate additional techniques that blend both ML and modern geostatistical simulation. The first of these is termed Post Simulation Prediction (Figure 9), where multivariate simulation of the predictor variables is first performed, reproducing features such as representative distributions (histogram) and continuity (variogram), which may be reliably calculated, parameterised, and targeted based on their relatively abundant data. The ML model is then applied on the block model scale to predict geometallurgical responses. In doing so, spatial structure is transferred to the geometallurgical variables, based on the realistic structure of the underlying predictors. No

spatial estimation or averaging of any sort is performed, thereby avoiding bias issues associated with nonlinear and nonadditive features of geomet responses.

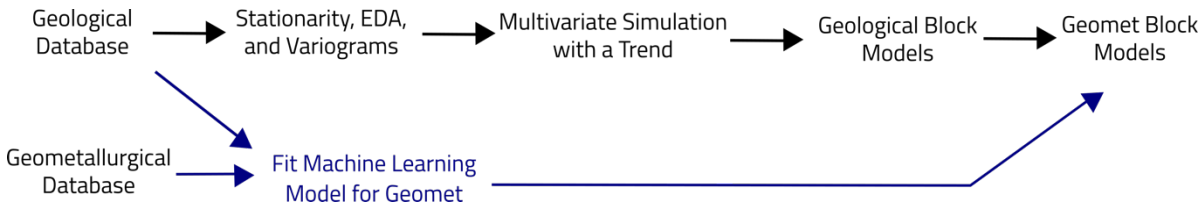

Figure 9. Prediction of geometallurgical response through applying a machine learned model to multivariate realisations of the predictors.

Figure 10 displays two realisations of predicted BWi from the porphyry example, where the BWi ML model (Figures 5 and 6) are applied to realisations of the categorical (alteration) and continuous predictor variables. Sericitic alteration is removed from Figure 10, so that uncertainty in the Potassic (central) and Propylitic (edges) is visualised with respect to the restricted BWi data (coloured tubes) and predictor data locations (grey lines). These realisations and the full set (generally dozens or hundreds) could be summarised to a single E-type or localised model, as mentioned above.

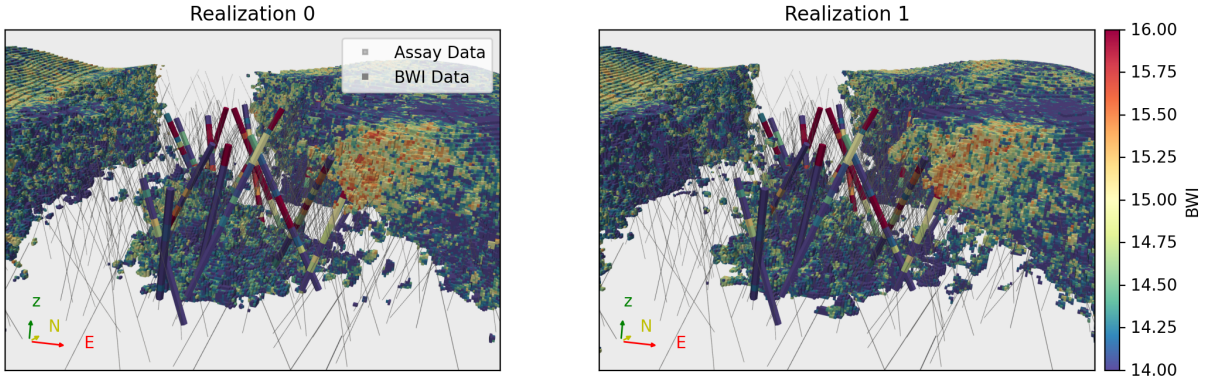

Figure 10. Realisations of BWi resulting from the Post Simulation Prediction approach.

Workflow 4: Direct Cosimulation Prediction

Recent advances in geostatistical simulation have contributed to the success and increasing use of Post Simulation Prediction, including the ability to characterise complex multivariate relationships of the predictor variables, as well as any non-stationary features (trends) that exist, following the transformation and imputation workflow that was previously described. Modern simulation software can harness the power of modern PCs or cloud computing, so that the simulation of dozens or hundreds of predictor variables is feasible. Programmatic based software allows for automation of many steps, reducing practitioner demand in the workflow setup or model updates. Nevertheless, as the number of predictor variables grows, the practitioner and computational demand will grow as well. Results for each variable should be carefully validated, with numerous parameters (e.g., variograms, histograms, and correlations) requiring validation and potential supervised tuning. Imputation methods exist for managing heterotopically sampled data but could become strained with more extreme cases.

These concerns may motivate Direct Cosimulation Prediction (Figure 11). After learning the predictor-response relationships in locations where the variables are collocated, the ML models are applied for predicting the geomet responses across all predictor data locations. The resulting geomet response will be more densely available, permitting assembly of parameters such as a variogram to permit its conventional estimation across the deposit. The resulting spatial model involves estimation, thereby raising concern associated with averaging of nonlinear and nonadditive geomet variables. Critically, this model is not used as the final model result. It is instead used as a secondary variable for conditioning cosimulation of the geomet variable. Only actual geomet data would be used as the

primary data within this cosimulation setup, so that no averaging related bias will be present in the final models. This workflow is practically attractive as it avoids the computational burden associated with simulating all predictor variables, while circumventing the need for imputing more extreme heterotopic predictor data.

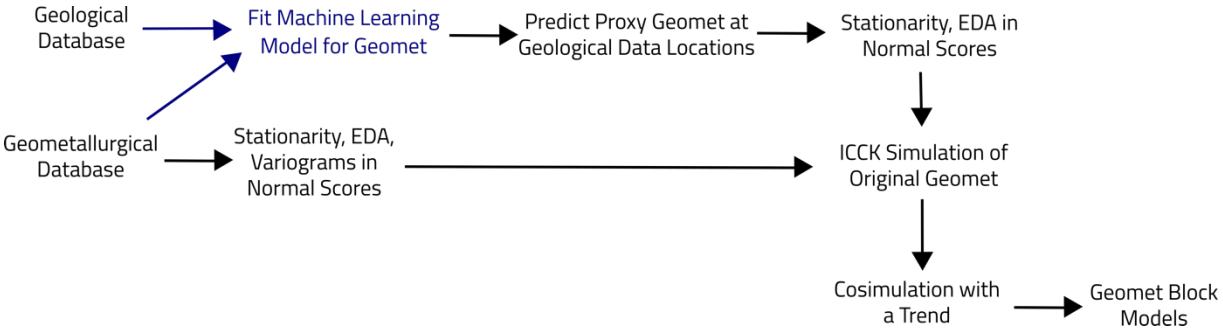

Figure 11. Prediction of geometallurgical attributes by machine learning at the data locations followed by cosimulation to generate geomet block model realisations.

Workflow 5: Bayesian Updating

There are alternative approaches that could be considered depending on the nature of the spatial distributions and relationships between attributes. Bayesian Updating (Figure 12) could be considered as an approach to merge predictions of the geomet attributes given geologic inputs and spatial predictions of the geomet attributes. The advantage of this approach is that distributions of uncertainty in the geomet variables is predicted directly, and simulation need not be considered.

The central feature of Bayesian Updating is to merge two conditional distributions coming from different sources into a posterior distribution. The procedure would be repeated for each geomet variable one at a time. This merger is done in normal score units, so the geomet variables must be normal score transformed. The first conditional distribution is of the geomet variable conditional to geological variables. The ML model is applied considering the geological block model to get the conditional mean. The conditional variance is taken from the k-fold validation performed as part of the ML prediction. The distributions are assumed Gaussian, which is reasonable given that everything is being computed in normal score units. The conditional variance could be increased away from geological conditioning data considering a relation to the kriging variance. The second conditional distribution comes from the local measurements of the geomet data. Simple kriging of the normal score transform of the geomet data provides a conditional mean and variance at each location based on the local geomet data.

The conditional distributions of the geomet variable given geology and the conditional distributions of the geomet variable given spatial information are combined with the Bayesian Updating methodology (Deutsch and Zanon, 2007). The result is a distribution of the normal score geomet variable at each location. A large number of quantiles could be back transformed from each local normal distribution to create a non-parametric distribution of the geomet variable at each location. This approach discretises the non-parametric distribution for further evaluation while preserving variability. These distributions could be summarised and post processed as appropriate. The result gives local uncertainty and avoids smoothing, but no simulated realisations are available for joint uncertainty assessment.

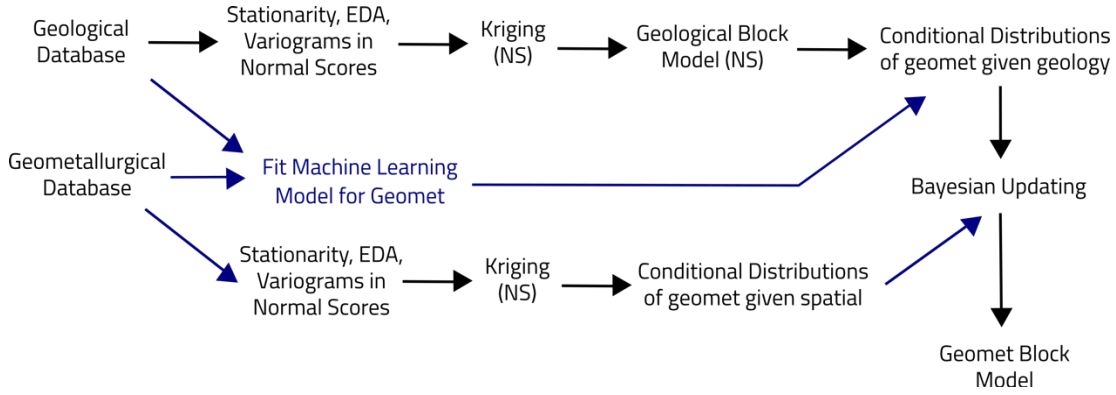

Figure 12. Workflow for Bayesian Updating to integrate geological and geometallurgical data with machine learning and geostatistics for a geomet block model.

Post-processing Considerations

These workflows generate either a single geomet block model or simulated realisations (geomet block models). A simulation approach defers averaging until the last possible instance, somewhat mitigating additivity challenges that are present with many geomet variables and providing options to better quantify the high and low values that can be expected when mining rather than smoother averages. Post-processing the multiple realisations to generate a single model for providing to mine planning operations may be completed with localisation or E-type (block-by-block) averaging. The application of these geomet models varies widely, from determining uncertain areas for further sampling campaigns to mine planning to maximise revenue and balance process feed requirements.

DISCUSSION

Increasingly, there are various data sources to consider. At a minimum we would have high resolution scanning data, conventional assays, and metallurgical test work. There may be multiple drill hole types, assay types, and a variety of geochemical and geomechanical measurements. Considering a specific target scale is critical. Extreme high-resolution data must be scaled to a larger scale. There will be inevitable mixing induced by blasting and loading at the mining face. Further mixing will take place in hoppers, recirculation, and introduction of additional material from stockpiles. In general, we would suggest simulation at the scale of the data, then application of an appropriate averaging schema.

There are pros and cons to the workflows described above. Selection of the most appropriate workflow depends on many factors including the amount and quality of the available geomet data, the amount and quality of geology data, the nature of the relationship between the geomet and geology variables, the purpose of the study, the time available for the study, and familiarity with the required software and tools. Simulation of the geomet variables using one of Workflow 2, 3 or 4 would be recommended in many situations. The reference data used for some of the figures provides an interesting opportunity to quantitatively compare and contrast the different workflows. This is outside the scope of this paper but provides an opportunity for future work.

CONCLUSION

The selection of where and when to inject ML within the geostatistical resource modelling workflow has been discussed. In general, it is used with a large variety of geochemical and grade variables to provide pre-posterior predictions at data locations. Then, geostatistical techniques are the priority to provide spatial predictions away from these data locations. There are many challenges: (1) geomet variables do not average linearly so smooth estimation and regression techniques are problematic, (2) there are relatively few geomet data locations exacerbating the problem with smooth estimation, (3) simulation solves the challenges with smoothing, but leads to potential difficulty managing multiple realisations. The workflows for practical geomet modelling are emerging.

REFERENCES

Aitchison, J., (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological*), 44(2), 139-160.

Barnett, R.M., Deutsch, C.V. (2015). Multivariate imputation of unequally sampled geological variables. *Mathematical Geosciences*, 47, 791-817.

Barnett, R.M., Manchuk, J.G., Deutsch, C.V. (2014). Projection pursuit multivariate transform, *Mathematical Geosciences*, 46, 337-359.

Boisvert, J.B., Deutsch, C.V. (2012). A short note on a general framework and software for localization of mining reserves. *Proceedings of the 14th Annual CCG Meeting*. Deutsch, C.V. and Boisvert, J.B. (eds.). Edmonton, Alberta, pp. (313) 1-4

Boisvert, J.B., Rossi, M.E., Ehrig, K., Deutsch, C.V. (2013). Geometallurgical modeling at Olympic Dam Mine, South Australia. *Mathematical Geosciences*, 45, 901–925.

Chen, T. & Guestrin, C., (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, pp. 785–794.

Deutsch, C. V. and Journel, A. G. (1998). GSLIB: A Geostatistical Software Library and User's Guide, 2nd edn. Oxford University Press - New York, New York.

Deutsch, J.L., Palmer, K., Deutsch, C.V., Szymanski, J., Etsell, T.H. (2016). Spatial modeling of geometallurgical properties: techniques and a case study. *Natural Resources Research*, 25, 161–181.

Deutsch, C.V. and Zanon, S.D. (2007). Direct prediction of reservoir performance with Bayesian updating. *Journal of Canadian Petroleum Technology*, 46(02).

Enders, C. (2010). *Applied Missing Data Analysis*. Guilford Press - New York, New York.

Friedman, J.H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38, 367-378

Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 15 pp.

Minniakhmetov, I. and Clarke, D. (2023). SBRE framework – application to Olympic Dam deposit: techniques and a case study. *Proceedings of Mineral Resource Estimation Conference 2023*. AusIMM, Perth Australia. 266–276.

OpenAI (2019). Deep Double Descent. https://openai.com/research/deep-double-descent [accessed 5 May 2023].

Pawlowsky-Glahn, V., and Olea, R. A. (2004). *Geostatistical Analysis of Compositional Data*. Oxford University Press – New York, New York.

Qu, J. and Deutsch, C.V. (2018). Geostatistical simulation with a trend using Gaussian mixture models, *Natural Resources Research*, 27(3), 347–363.

Scott, T.R., Gous, C., Muller, J., and Deutsch, J.L. (2023). Applying machine learning to predict the impact of changing iron ore properties on blast furnace performance, *SAIMM Geomet 2023*.

Silva, D.S.F and Deutsch, C.V. (2016) Multivariate data imputation using Gaussian mixture models, Spatial Statistics, 27, 74-90.

Dr. Jared Deutsch

President
Resource Modeling Solutions Ltd.

Dr. Deutsch leads Resource Modeling Solutions and directs geometallurgical modeling projects. Dr. Deutsch has experience in mineral resource estimation and geometallurgical modeling for complex hard rock deposits. He specializes in integrating metallurgical data with machine learning into mineral resource models for better process predictions and increased mine value. He holds advanced degrees in both extractive metallurgy and geostatistics. Prior to joining Resource Modeling Solutions, he was a geometallurgist and geostatistician on the Carlin Trend at Newmont. Dr. Deutsch is the founder and editor of Geostatistics Lessons providing guidance on the application of geostatistics.