Grade control with Ensembled ML: A comparative case study at the Carmen de Andacollo copper mine

Camilla Zacche da Silva, Jed Nisenson, Jeff Boisvert

Ore/Waste delineation is important in mining projects as it is linked to revenue, ore misclassification impacts profitability. The main goal with grade control is to predict the destination of extracted material based on all available data collected. Common approaches for grade control are based on estimated maps obtained through kriging, inverse distance estimation or nearest neighbor methods. Nevertheless, traditional geostatistical frameworks require first and second order stationarity inside a defined domain. Machine learning algorithms provide flexibility and simplicity when integrating data and incorporating complex patterns that cannot be easily accounted for with linear geostatistical workflows. The methodology implemented uses machine learning algorithms to account for non-stationarity throughout the area of interest. Two methods are used: Elliptical Radial Basis Function Network and Support Vector Regression; these methods are applied to obtain trends that are incorporated in a collocated co-kriging framework to generate a final grade model while account for non-stationarity. This workflow allows for automatic data incorporation and can be used to improve grade control decision-making. The workflow is demonstrated on Teck Resources Limited's Carmen de Andacollo copper mine in Chile, where samples from 10 blasts are used for validation. Ordinary kriging, inverse distance and a simulation-based approach are also used to model the 10 blasts and each model is evaluated according to a benchmark. The trends generated with ML algorithms, when incorporated to the geostatistical framework have demonstrated in this case study improvement over the traditional approaches. Support vector regression obtained a reduction in misclassified material of 8% relative to inverse distance, 1.12% relative to ordinary kriging and 1.16% relative to the simulation-based approach. In the case of elliptical radial basis functions, the models obtained a decrease in misclassified material of 12%, 5.4% and 5.7% compared to inverse distance, ordinary kriging and simulation-based approaches, respectively.

Introduction

Ore/Waste delineation is fundamentally important in mining projects as misclassifications directly impact project value (Verly, 2005). Grade control involves assigning material destinations based on all available data, including blast holes, drill holes, remote sensing, etc. The most widely used approaches involve assigning an economical cutoff grade based on kriging (Matheron, 1963) or inverse distance (ID) models. However, geological data is variable at all scales due to the inherent heterogeneity resulting from the deposition process combined with sparse data collection, which leads to uncertainty when predicting values at unsampled locations. Estimation through traditional geostatistical frameworks requires the assumption of first and second order stationarity inside a defined domain. Most geological data are featured with location dependency, therefore, a non-stationary behavior.

By contrast, machine learning (ML) provides a set of techniques that can automate the process of analytical model building, based on the premise that systems can learn from the data. In this framework, data sets that are representative of the targeting problem can be used to learn and exploit meaningful relationships from the data without the assumption of stationarity.

The use of ML has been increasingly applied to resource modeling problems (Dowd and Saraç, 1994; Kapagerdis, 1999; Tahmasebi and Herzakahni, 2012; Dai et al., 2014; Gangappa et al., 2017, Maniar et al., 2018, Tomislav, 2018, Samson and Deutsch, 2018; Samson and Deutsch, 2019 Walch et al., 2019). However, ML does not guarantee data reproduction at sampled locations, as geostatistical approaches do, and it does not explicitly account for spatial correlation.

Extensive research has been dedicated into characterizing the non-stationarity of geological settings by incorporating a trend. So, the final predicted model is accurately representative of the phenomenon (Journel and Huijbregts, 1978; Deutsch and Journel, 1998; Lloyd, 2011; Qu, 2018). Trend is understood as large-scale spatial varying feature of the attribute of interest and can be modeled manually or computationally (Qu, 2018). The present paper focus on use ensembled ML methods to build a trend model to be incorporated on grade control prediction.

As previously mentioned, ML has been increasingly applied to resource modeling problems. Some examples are the work of Li et al. (2011) which applied random forests to predict the mud content on an Australian margin as well as the approach developed by Tomislav et al. (2018) on spatial random forests by incorporating Euclidean distances to calculate buffer zones. ML algorithms are flexible and can capture non-linear relationships on the data. Nonetheless, differently from the geostatistical analogue, ML predictions do not guarantee data reproduction at the data locations. Thus, Samson and Deutsch (2018; 2019) have developed a hybrid model that incorporates the predictions of elliptical radial basis function network (ERBFN) to the collocated cokriging (CCok) framework to improve accuracy on estimated models. The menu of ML techniques is wide and such workflow can be explored with other learning algorithms. Thus, in the present paper, support vector regression (SVR) will also be explored as an alternative in obtaining the trend model to be incorporated in the estimation process, which obtained very promissing results.

Thus, the methodology porposed by Samson (2019) will be applied in the context of grade control, aiming to enhance accuracy of estimated values and, therefore, better decision making relative to the extracted material. The case study herein presented, makes a comparative evaluation of two different ML techniques applied to predicting the destination of the mined material from 10 different blast areas at the Carmen de Andacollo copper mine.

To evaluate the applicability of each ML approach, results will be confronted with the currently used approach in geostatistics: ordinary kriging (OK) and inverse distance (ID) estimation. Also, a simulation-based approach, *intelligent grade control (IGC)* (Vasylchuck and Deutsch, 2017) will be run and evaluated as well. Realizations of the geological reality were generated to serve as a benchmark and assess model performance.

This paper is outlined as follows: the background section describes machine learning methodologies applied on the case study, along with a brief review on ensemble methods. The methodology section presents a detailed description of the workflow proposed and of the evaluation procedure of the methodologies. A results section is presented with the performance discussion of each methodology, followed by the conclusions.

Background

RBFN/ERBFN

Radial Basis Functions neural networks (RBFN) are fundamentally different from the popular multilayer perceptron, which have many layers and the nonlinearities are introduced by the repeated compositions of activation functions on different hidden layers. The RBF has only one hidden layer, with a specific activation function type that introduces nonlinearities to the system (Aggarwal, 2018). The structure of the RBFN is as follows:

- The input layer, which transmits the input features to the hidden layer. The input layer does not perform any transformation or computation on the input data.

- The hidden layer consists of nodes that contain the activation function, that in turn, is defined by μ_i and σ_i. Although μ_i is specific to each activation function, the σ_i is often defined as the same value for all nodes (Aggarwal, 2018). Also, the total number of nodes in the hidden layer is set to be less than the total number of training points in the input layer. This is an important hyperparameter in the system.
- The weights w_i are assigned to the values from the hidden layer to the output layer to generate a prediction. The weights are a dynamic quantity, that changes as the learning process from the RBFN occurs.
- The output layer produces the output values through a weighted sum of the hidden layer values.

The present study made use of the algorithm developed by Samson and Deutsch (2019), to generate a trend model. In this framework, the activation function considered in the NN is Gaussian described by equation 7:

$$G(x_i) = e^{-(\sigma(\mu_i - x_i))^2}$$

Where σ is the radii or spread of the Gaussian activation function, μ_i is the center of the Gaussian and x_i is the *i*th training point. The location for each Gaussian center is optimized through K-means clustering algorithm (Loyd, 1957; Forgy, 1965; Samson and Deustch, 2019; Samson, 2019), which divides the data into *m* clusters with equal variance (Pedregosa et al., 2011; Samson and Deutsch, 2019; Samson, 2019). For that, the data is partitioned while minimizing the moment of inertia, which is defined by equation 8 (Samson, 2019):

$$I = \sum_{i=1}^{n} \min(\|x_i - \mu_j\|)^2$$

The term $\|x_i - \mu_j\|$ is the Euclidean distance between *i*th point to the cluster center it is assigned to. Then, the cluster centers are adjusted according to equation 9:

$$\mu_j = \frac{1}{m_j} \sum_{j=1}^{m} x_i$$

m_j is the *j*th cluster. Equations 2 and 3 are calculated iteratively until a potential minimum of inertia is reached between all clusters. The algorithm is initialized with random center locations (Samson, 2019).
Considering equation 7 as the activation function in the hidden layer, this approach aims at generating a *N*-dimensional surface from which the estimated outputs are obtained (Samson, 2019), $\mu_i - x_i$ refers to the Euclidean distance between the cluster center and the estimated point. However, Samson (2019) proposes the use of Mahalonobis distance instead of Euclidean, aiming to incorporate anisotropy to the Gaussian function. The Mahalonobis distance is given by equation 10.

$$M_d = \sqrt{d^T S^{-1} d}$$

S^{-1} is the inverse covariance matrix between the cluster center and the estimation point (Todeschini et al., 2013; Samson 2019), and *d* is the difference between the estimation point and the cluster center. Given that the gaussian uses the Mahalonis distance, the network is referred to as elliptical radial basis function (ERBFN) and have the same structure as the RBFN, their distinction being on the distance term. Samson and Deutsch (2019) have shown that this shift on the distance function greatly improves data spatial pattern reproduction by the algorithm.

The output layer computes a weighted combination of the hidden layer values which leads to the predicted value. The predicted values are continually evaluated during training by computing the error of each prediction. According to the error obtained at each step the weights in the network are adjusted. To initialize the process, the user must set the number of nodes contained in the hidden layer. That is, the number of units that will process the input features. The number of nodes directly affects the ERBFN

performance, given that too few nodes lead to underfitting and too many nodes increase computational time and lead to overfitting.

SVR models

Support vector machines (SVM) introduced by Vapnik, (1995, 1998) solves a classification problem. However, the concept can be extended to a regression context by introducing an ε-intensive region around the fitted function (Awad and Khanna, 2015). The goal is to obtain $f(x)$ that has a maximum of ε deviation from the actual observed target values while retaining the regression coefficients, w, as small as possible. Considering a training dataset x_n with N observations and response values y_n, the aim is to find the linear function:

$$f(x) = \langle w, x \rangle + b$$

In this context, small w values are obtained by means of minimizing the Euclidean norm $\frac{1}{2}\|w\|^2$. Formally, it can be stated as a convex optimization problem:

$$\min \frac{1}{2}\|w\|^2$$

Subject to the constraints:

$$\begin{cases} y_i - \langle w, x \rangle - b \leq \varepsilon \\ \langle w, x \rangle + b - y_i \leq \varepsilon \end{cases}$$

The optimization presented in equations 1 and 2 is feasible when exists a function f that approximated the pair (x, y) with ε precision. However, there is not always such function. When the solution is not feasible some errors are allowed by means of slack variables, ξ_n and ξ_n^*

While ensuring that it is as flat as possible, for that, expression 2 must be minimized:

$$\frac{1}{2}\|w\|^2 + C\sum_{n=1}^{N}(\xi_n + \xi_n^*)$$

Subject to the constraints:

$$\begin{cases} \forall n: y_n - \langle w, x \rangle \leq \varepsilon + \xi_n \\ \forall n: \langle w, x \rangle - y_n \leq \varepsilon + \xi_n^* \\ \forall n: \xi_n^* \geq 0 \\ \forall n: \xi_n \geq 0 \end{cases}$$

The constant C is a positive value that controls the penalty imposed on values that lie outside the ε-intensive region, that is, a regularization factor. ξ_n and ξ_n^* allow error to exist, analogue to the soft margin concept on the support vector machine formulation. Obtain the solution for nonlinear problems Lagrange multipliers are introduced, which leads to the following optimization problem:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)\langle x_{i,}x_j\rangle + \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i)$$

Subject to the constraints:

$$\begin{cases} \sum_{n=1}^{N}(\alpha_n - \alpha_n^*) = 0 \\ \forall n: 0 \leq \alpha_n \leq C \\ 0 \geq \alpha_n^* \geq C \end{cases}$$

The function to predict new values is described in equation 4:

$$f(x) = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*)\langle x_n, x\rangle + b$$

And the conditions Karush-Khun-Tucker are necessary to obtain optimal solutions.

$$\begin{cases} \forall\, n\colon \alpha_n(\varepsilon + \xi_n^* + y_n - \langle w, x_n\rangle - b) = 0 \\ \forall\, n\colon \alpha_n(\varepsilon + \xi_n - y_n + \langle w, x_n\rangle + b) = 0 \\ \forall\, n\colon \xi_n(C - \alpha_n) = 0 \\ \forall\, n\colon \xi_n^*(C - \alpha_n^*) = 0 \end{cases}$$

However, some solutions are not obtained with a linear function. This is overcome by replacing the dot product $\langle x_{i,} x_j\rangle$ in equation 3 with a non-linear kernel function $G(x_i, x_j) = \langle \varphi(x_i) | \varphi(x_j)\rangle$ which maps the data into a high dimensional space. Therefore equation 3 becomes:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)G(x_i, x_j) + \varepsilon\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) - \sum_{i=1}^{N} y_i(\alpha_i + \alpha_i^*)$$

And the function to predict new values is:

$$f(x) = \sum_{n=1}^{N}(\alpha_n - \alpha_n^*)G(x_{n,} x) + b$$

Most machine learning packages for support vector regression come with built in different kernels such as linear, polynomial and gaussian. So, opposed to traditional regression methods that aim at minimizing the deviation from predicted values and observed values, SVR aims at keeping the error below a threshold.

Collocated Cokriging

Exhaustively available secondary information is often incorporated through the collocated cokriging framework (Xu et al., 1992), which retain solely the secondary information at the location being estimated. This approach is well known as the Markov-type formulation (Almeida, 1993; Almeida and Journel, 1994; Journel, 1999), and has become popular due to its computational efficiency and simplicity when compared to the full cokriging system, hence its use on the present case study.

Methodology

The proposed methodology for ore/waste delineation follows 5 overall steps. Each step will be expanded upon below.

Step 1: collect all data available;
Step 2: train algorithm on available data (ML framework);
Step 3: generate spatial grade model of the variables of interest;
Step 4: apply economical cutoff grade to determine material destination;
Step 5: generate minable plan using available dig limit algorithms.

Data set

The proposed methodology is demonstrated on 10 blast areas in the Teck Recourses Limited, Carmen de Andacollo (CDA) copper mine. The mine is located in the coastal range of central Chile, adjacent to the town of Andacollo. The cutoff grade considered is the breakeven value of 0.2% of total Copper. The data considered to build each estimated model consists of the blast holes specific to each area. The data is nearly isotopically sampled at a spacing of 5m with an average of 120 blast holes per area. Table 1 presents the statistical summary for each blast area in the case study.

Table 1: Statistical summary for data collected over ten blast areas at the Carmen da Andacollo copper mine.

Blast Number	Number of samples	Mean	Std. Deviation	CV	Maximum	Minimum
1	133	0.31	0.17	0.54	1.06	0.03
2	106	0.26	0.08	0.3	0.53	0.05
3	115	0.50	0.2	0.40	0.94	0.13
4	123	0.40	0.14	0.34	0.74	0.15
5	170	0.41	0.11	0.28	0.81	0.15
6	143	0.21	0.11	0.52	1.04	0.07
7	96	0.21	0.1	0.47	0.86	0.07
8	117	0.22	0.24	1.08	1.54	0.02
9	120	0.24	0.15	0.61	1.09	0.07
10	124	0.27	0.45	1.64	4.76	0.04

To tackle the case study in a semi-automated framework, aspects of the training (when using ML) and modeling process are performed in a semi-automated fashion. In the case of ML, hyperparameters needed for training and fit of an optimal model are not defined problem specific. In the geostatistical context, variogram models are built automatically and fine-tuned manually when needed.

It is noted from Table 1 that each blast area has distinct statistical features. For example, Blast 2 presents $CV = 0.3$ while blast 10 $CV = 1.64$ with a large range in minimum and maximum grades. Outliers are detected on the data sets and are capped on the P90 values of each set. The varying range of complexity and on the data sets surely affects the training process for the ML algorithms as the selection of hyperparameters is not defined problem specific.

Grade model

For the case study presented, five different grade models were generated and evaluated. Two of the grade models incorporated trends obtained through machine learning techniques as previously discussed. The remaining three were generated through ID, OK and IGC respectively. The process to obtain each estimated grade model is outlined in the following sections.

Inverse distance

Inverse distance is one of the conventional methods used to build grade control models given its simplicity to implement. ID estimation is based on calculating weights for the samples given the distance to the estimation location. The following equation presents the ID estimator:

$$z^* = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i}$$

Where z^* denotes the estimated values at an unsampled location, z_i is a sample value and w_i are the weights assigned to each sample given that $w_i \propto \frac{1}{d^\omega}$. d is the Euclidean distance between a sample and the estimation location. The model can be modified by increasing the power on the distance term. Higher ω lead to less smooth models. The most common value used is $\omega = 2$.

Ordinary Kriging

The basis for ordinary kriging is to build a weighted linear combination of the available samples in such a way as to minimize the error variance, subject to the unbiasedness constrain: $E\{z^* - z\} = 0$. The weights obtained in ordinary kriging directly depend on the choice of a variogram model for the data set. This enables the kriging algorithm to obtain insight on the anisotropy pattern of the geological phenomenon. Obtaining a variogram model for the kriging framework is a prerequisite that can be performed manually by the user or automatically, through numerical algorithms.
Another important aspect of kriging is the sample search strategy, which controls the samples that are to be included in the estimation process. Given that the blast hole data set is regularly spaced, the sample search is kept the same for all the blasts and is defined as shown in Table 2:

Table 2: Sample search strategy for the OK estimated model

Minimum number of samples	Maximum number of samples	Anisotropy	Search radius
2	16	Variogram determined	30m

Intelligent Grade Control

Intelligent Grade Control (IGC) is an unsupervised spatial prediction algorithm based on local multivariate simulation (Vasylchuck and Deutsch, 2018; 2019). The algorithm calculates the expected profit at the grid resolution, specified by the user. The expected profit is used to determine the waste and ore inside the area defined and used to obtain optimal dig limits for material extraction. Ideally, blast movement should be considered, however, this information is not currently available. The algorithm is completely automated, in which the only parameter set by the user the cutoff grade value used and the grid resolution. For the current case study, the grid resolution defined is the same for all grade models evaluated. The cell size is defined as 1.25mx1.25m in the XY directions and 0.25m vertically. The number of nodes in each grid is dependent on the area covered. Also, the cutoff grade value is the same in all cases analyzed, accordingly to the company's determination of 0.2% Total Copper.

Machine Learning combined with Collocated cokriging models

To generate the grade control models where ML is applied, the first step is to train the ML algorithm. This procedure demand that hyperparameter setting for each technique be specified prior to the training stage. For example, on ERBFN algorithm the user must define how many nodes will constitute the hidden layer in the network. The number of nodes in the network can viewed as the degrees of freedom of the model, as

having too many nodes often lead to overfitting (Samson, 2019) and can be computationally inefficient. In the SVR framework, the user must specify the type kernel used, the regularization factor, C and the gamma value, in case of a RBF kernel. As discussed in the background section, the C hyperparameter controls the penalty of data the lie outside the ε- intensive region, ergo, directly affecting the final model generalization capability. Larger C values impose higher penalties and lower C values impose lower penalties. It is important to find a balance between the training data fit and model generalization. Another important parameter in the SVR framework, given that the kernel is a RBF is the gamma hyperparameter. This controls the influence radius of the function. High gamma values lead to functions that decrease rapidly and smaller gamma values lead to functions that decrease more slowly, therefore, have higher range of influence. Nonetheless, for both techniques, these hyperparameters are intimately linked to the data complexity and obtaining optimal values is challenge on its own. So, to alleviate possible over and underfitting, ensembled techniques are used.

ERBFN Training, Ensemble and Grade Model

As discussed in the introduction section, the ERBFN algorithm (Samson, 2019) optimizes most of the parameters considered in the training phase, except the number of nodes of the hidden layer and the learning rate. In thiscase study the number of nodes is defined as a proportion of samples on each blast area given that the approach is proposed in a semi-automated fashion, the network is not tuned specifically to each blast area. Figure 1 shows the average R^2 over the 10 databases considered from predictions obtained from varying number of nodes in the ERBFN in a 5-fold cross validation set.

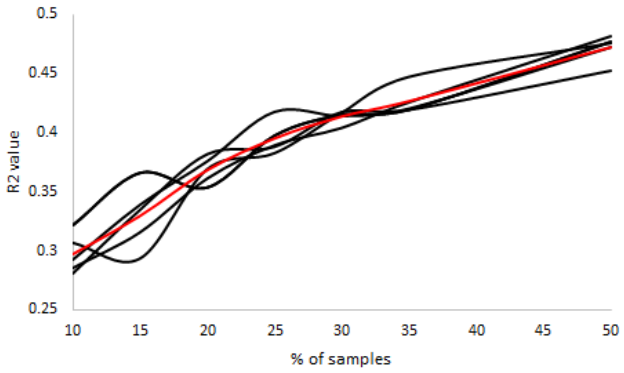

Figure 1: R^2 values obtained between predicted maps from ERBFN and data samples with varying number of nodes in the hidden layer (black lines represent each fold; red line represents the average over the folds)

It can be seen from Figure 1 that the average behavior of the R^2 value is significantly less stable considering as the number of nodes in the hidden layer less than 25% of points in the training set. Also, it presents a consistent increase as the proportion of training points is higher. So, the set of nodes considered varies between 30% of the training points up to 45% of the training points. Once the set of nodes, $S_n\{n_1, n_2, \dots n_n\}$, for the hidden layers is defined the ERBFN algorithm trains and fit a different network for each n_i node values. Each network will generate a prediction. Finally, the different predictions are combined through averaging.

The trend models obtained through ERBFN algorithm are presented in Figure 2 along with the sample location map.

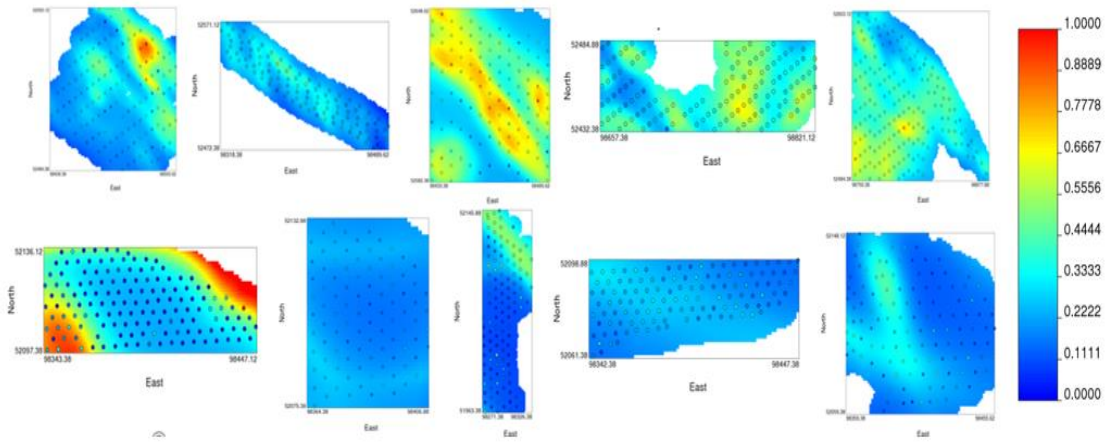

Figure 2: Trend models obtained through ERBFN algorithm with sample data set on 10 blasts

To generate the grade control model, the trend obtained through ERBFN algorithm are used as an exhaustive secondary variable on an intrinsic collocated cokriging framework. For that, the variogram models used previously on the OK approach are used in this framework as well. The cross variogram needed for the collocated cokriging is obtained from the correlation coefficient from each trend model to the sample data. The correlation values obtained are presented in Table 3.

Table 3: Correlation between trend and sampled values

Data	Blast 1	Blast 2	Blast 3	Blast 4	Blast 5	Blast 6	Blast 7	Blast 8	Blast 9	Blast 10
ρ	0.89	0.93	0.9	0.94	0.94	0.6	0.7	0.84	0.6	0.7

As well as kriging the sample search strategy must be defined. In this case the search is kept as similar as possible to the one used on the OK framework, however, as this is a collocated cokriging approach is also kept on the search the information from the trend model that is collocated with the node being estimated. Table 4 summarizes the search strategy applied for the ERBFN with collocated cokriging approach:

Table 4: sample search strategy applied to collocated cokriging approach

Minimum number of samples	Maximum number of samples	Maximum secondary (trend) samples	Anisotropy	Search radius
2	16	1	Variogram determined	30m

SVR Training, Ensemble and Grade Model

In the present case study, SVR is ensembled through stack generalization (Wolpert, 1992). This is a general method of combining multiple models to achieve greater predictive accuracy. The process consists in splitting the data set into training, validation, and test subsets. Then, different base models are defined. The base models generate predictions for the validation set and the pairs of predicted values and observed values are fed into a meta model which trains over these pairs. Then, the meta model generates the final prediction.

As previously mentioned, to fit an SVR model, previously to the training phases it must be defined what type of kernel will be used, the regularization parameter, C, and the in the case of RBF kernels, gamma. To build the stacked generalization two SVR with gaussian kernels are used with the following parameter setting: $C = 1$ and gamma=10; $C = 100$ and gamma=1. The meta model which is trained with the

predicted values from the base models is a linear kernel with $C = 1$. The ε-region accepted for the base models and for the meta model is $\varepsilon = 0.1$. The trend models obtained from the stacked generalization along with the sample data location map is show in Figure 3.

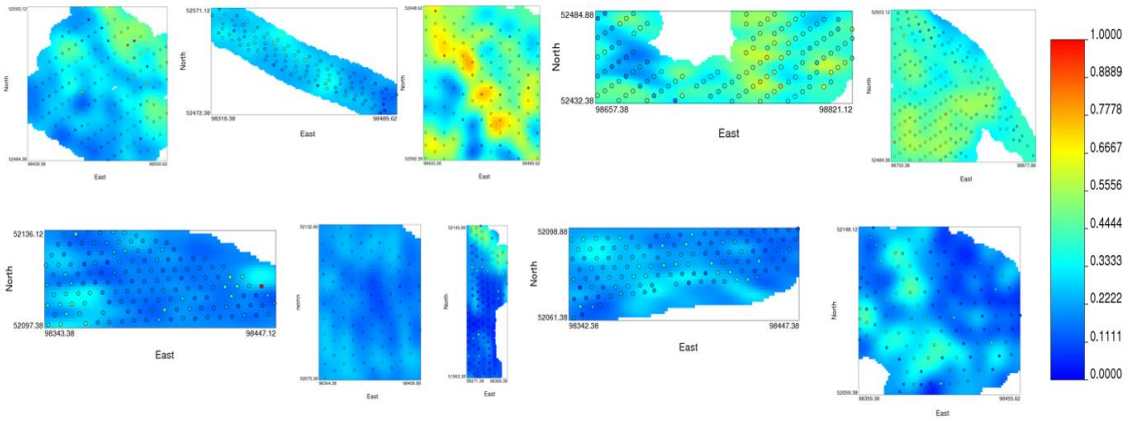

Figure 3: Trend models obtained through ERBFN algorithm with sample data set on 10 blasts

To generate the grade control estimated model, the trends obtained through the staked SVR algorithm are used in the collocated cokriging framework. As it is done with the ERBFN algorithm, in this case the variogram models used previously on the OK model are applied as well. The cross variogram needed for the collocated cokriging is obtained from the correlation coefficient from each trend model to the sample data. The correlation values obtained are presented in Table 5

Table 5: correlation between trend and sampled values

Data	Blast 1	Blast 2	Blast 3	Blast 4	Blast 5	Blast 6	Blast 7	Blast 8	Blast 9	Blast 10
ρ	0.63	0.74	0.86	0.76	0.96	0.57	0.78	0.77	0.52	0.62

Also, in the case for SVR combined with collocated cokriging, the sample search strategy is kept the same as the one applied on the ERBFN context presented in Table 4.

Economical cutoff and material destination

Given the grade estimated models the next step on the proposed workflow is to apply the economical cutoff grade to determine the material destination in each scenario. The company defines the material destination based on the break-even cutoff grade value. In this case total copper greater than 0.2% is considered ore and less than 0.2% is considered waste.

Optimal minable dig limits

There are a number of approaches possible to process minable dig limits, such as the use of grade control polygons. In the present paper the dig limits are obtained through an automated algorithm (Vasylchuck, 2018) that approximates the optimal dig limit in a two-step process. First, the locations are coded as the destination according to the cutoff grade and profit. Narrow zones of different destinations are considered problematic, such zones are visited on the second step of the algorithm which are enforced to modify the

destination of the material considering the profit. For further details on the algorithm reader can refer to (Vasylchuck, 2018).

Reference Models

The true reconcilation data is not available, the models built for this case study have the performance ranked based on the comparison simulated scenarios for each blast area. The simulated "truths" are obtained by means of Sequential Gaussian Simulation (Isaaks, 1990). For each blast hole data, 50 realizations are generated and one of the realizations is randomly drawn as the "true" distribution for the respective blast area. The simulated models are verified in terms of variogram reproduction as well as histogram reproduction. Then, the randomly drawn "truth" is processed to obtain the material destination for each blast as well as the optimal dig limits.

Results and Discussion

A fundamental part of grade control is the estimated model for the blast area. In the previous sections its been outlined how each estimated model is built. Prior to proceeding in processing dig limits and verifying the extracted material, the estimated models were evaluated through 5-fold cross validation. The mean squared error (MSER) is used as a metric of performance for each model over the ten blasts considered. Figure 4 presents the results obtained.

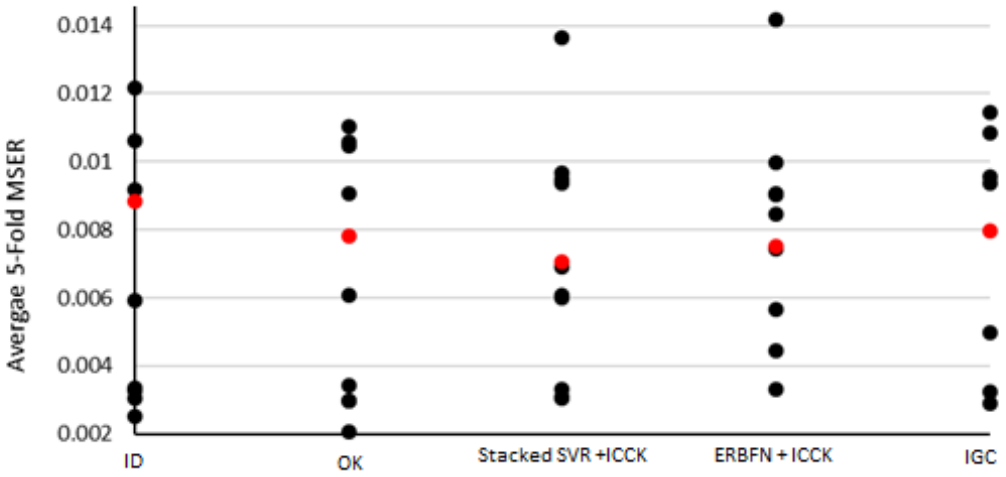

Figure 4: Mean Squared error obtained from 5-fold cross validation for the 10 blasts with each methodology applied

It can be seen from Figure 4 that ID estimation obtained the highest MSER overall blast areas, the average is $MSER_{ID} = 0.0088$ while the average values for Stacked SVR with CCok is $MSER_{SVR+CCok} = 0.0070$, approximately 25% lower than the one obtained with ID modeling. Also, ERBFN with CCok obtained better performance than ID, OK and IGC with $MSER_{ERBFN+CCok} = 0.0075$, with a decrease of 14% relative to ID average MSER. Stacked SVR also presented a decrease of the $MSER$ relative to OK and IGC of 10% and 11.5% respectively. On the other hand. ERBFN obtained a decrease on the average $MSER$ of 3.8% relative to OK and 5% relative to IGC.

Given the estimated grade models the material destination is defined accordingly to the cutoff grade as previously mentioned. Each block in the estimated grade model is declared as ore or waste. Ore blocks are coded as 1 and waste blocks are coded as 2, as presented in Figure 5.

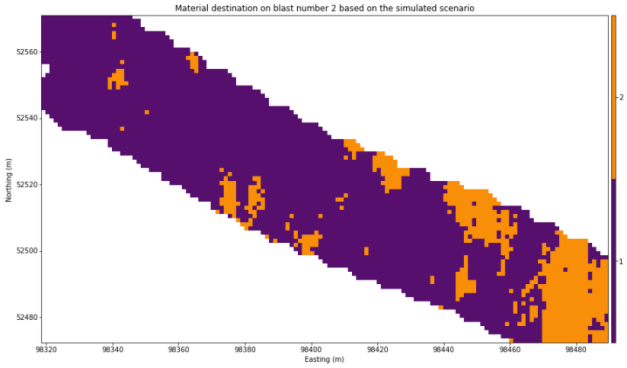

Figure 5: Material destination defined according to block value. Code 1 (purple) represents blocks declared as ore and code 2 (orange) represents blocks declared as waste inside the blast area. The map refers to blast area number 2, obtained from the simulated "truth"

It is noted from Figure 5 that defining the blocks as waste and ore is not sufficient to obtain a minable dig limit. For that, the models with the ore/waste classification are fed into the algorithm igc-dl.exe (Vasylchuk and Deustch, 2018). The algorithm calculates the optimal minable dig limit accordingly to the profit calculated for each block and the cutoff grade. Which leads to the corresponding map in Figure 6.

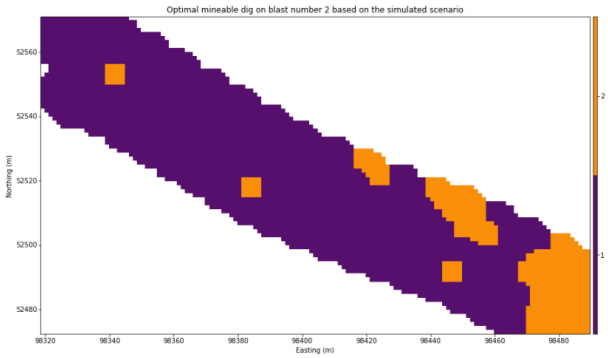

Figure 6: Optimal dig limit defined according profit on each block. Code 1 (purple) represents blocks declared as ore and code 2 (orange) represents blocks declared as waste inside the blast area. The map refers to blast area number 2, obtained from the simulated "truth".

Figure 6 shows the optimal dig limit defined for blast number 2 based on the profit and cutoff grade calculated from the simulated "truth". This process is performed for all 10 blast areas over the grade estimated models obtained through ID, OK, IGC, CCok with ERBFN and CCok with SVR. The total of ore/waste mined on each area is compared to that obtained on the simulated scenarios Following, the results obtained after the optimal minable dig limit for each method are presented. The amount of incorrectly classified material is based on the simulated "truths" of each blast area. Figure 7 presents the proportion of ore incorrectly mined, that is, which should have been mined as ore and is in fact mined as waste (type II error), over the 10 blasts for each methodology.

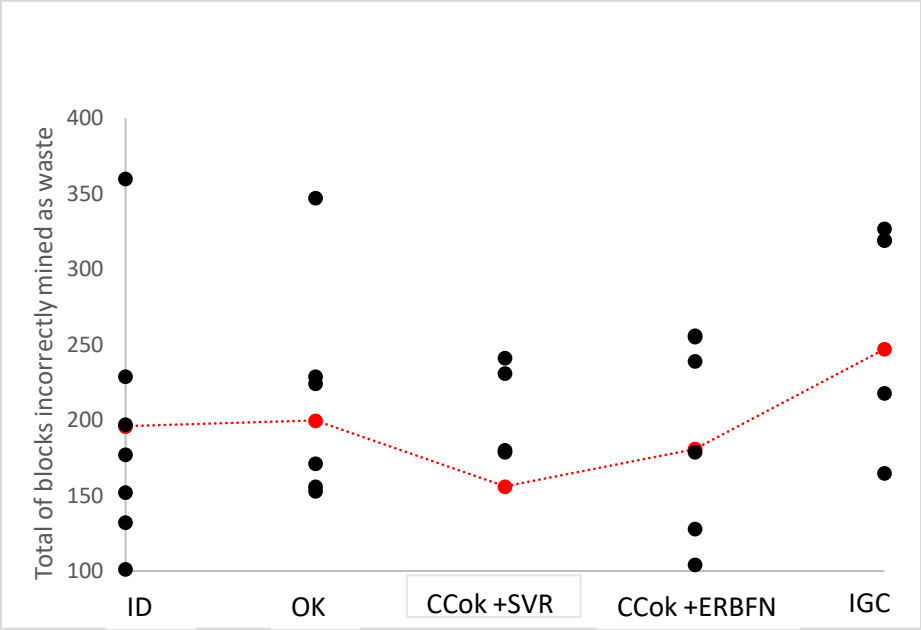

Figure 7 Total of blocks incorrectly mined as waste over 10 blast areas

It is seen from Figure 7 that the IGC obtained higher average error. IGC in this case, classified ore as waste, which indicates an underestimation of the values inside the area of interest. The capping applied on the data set may have affected IGC capability to predict higher grades. On the other hand, SVR with CCok obtained less type II errors than ID, OK, ERBFN with CCok and IGC. SVR with CCok obtained a reduction on type II error of 20% relative to ID and OK; and 36% relative to IGC. Figure 7 presents the type I error obtained.

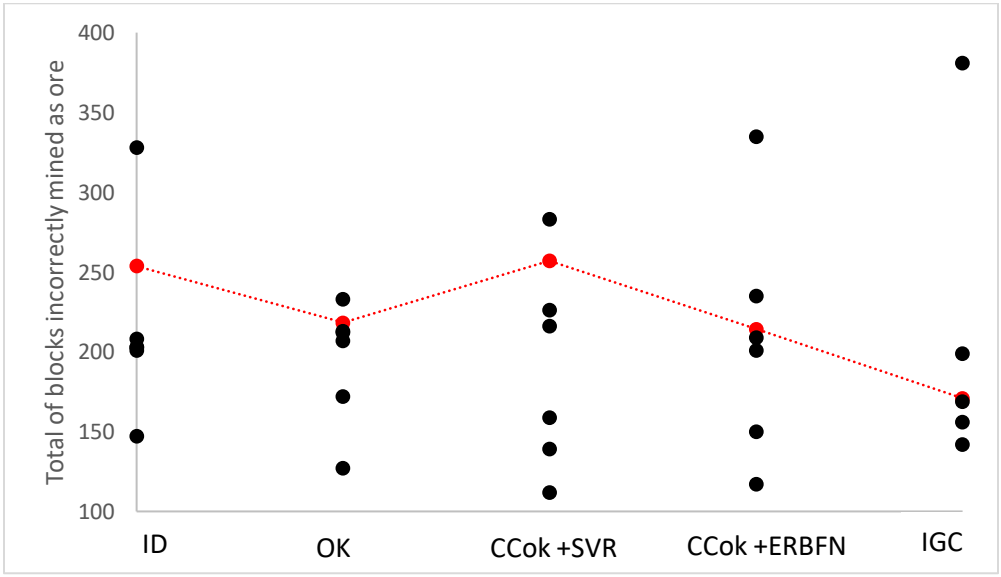

Figure 8 – Total of blocks incorrectly mined as ore over 10 blast areas

In contrast to Figure 7, IGC misclassified less waste than the other methodologies, with an improvement of 32% relative to ID and 21%, 33% and 20% less misclassified material relative to OK, SVR with CCok and ERBFN with CCok, respectively. ERBFN with CCOK, however, misclassified less material than ID, OK and SVR with CCok. The IGC underestimation of grade observed on Figure 7 reflected on a higher predictability of waste inside the blast areas. On the other hand, CCok with SVR presented the opposite behavior, not capturing adequately the lowest grades inside the blast areas. The results from Figure 7 and Figure 8 are combined and presented as the total amount of misclassified material over the ten blasts considered. This result is shown in Figure 9.

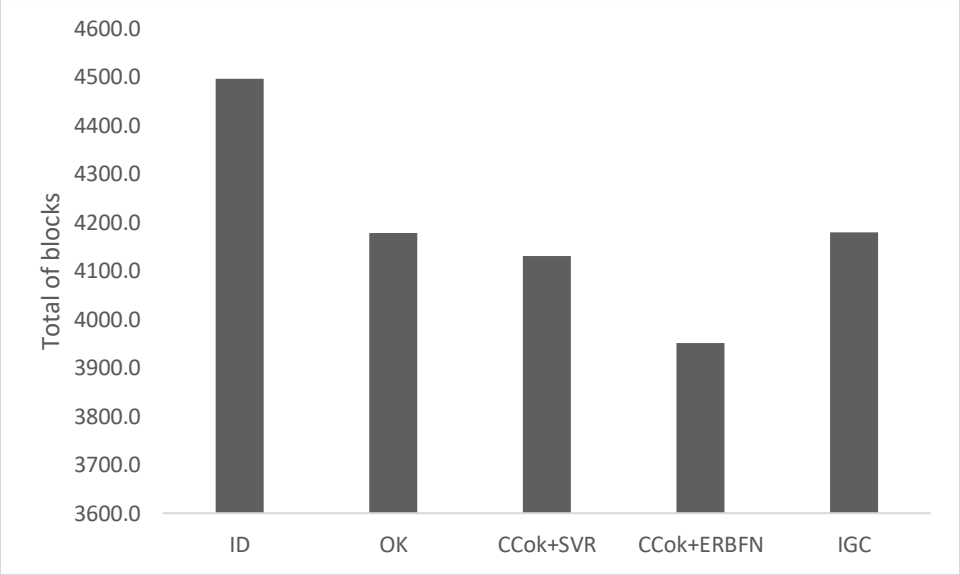

Figure 9- Total number of blocks incorrectly mined over10 blasts

Figure 9 shows that, CCok with ERBFN present best performance in this case study. CCok with ERBFN obtained a reduction in misclassified material of 12% when compared to ID; 5.4% less than OK, 5.7% relative to IGC and 4.3% relative to CCok with SVR. Even though SVR with CCok is outperformed by ERBFN it obtained a reduction of 8% in misclassified material relative to ID and 1.12% and 1.16% to OK and IGC respectively.
Often ML techniques require fine tuning of the hyperparameter during training phase, however, in the case study presented the hyperparameter setting was not chosen problem specific. The ERBFN set the number of nodes of the hidden layer as dependent on the number of samples of each blast areas as means to obtain a workflow as straightforward as possible. While SVR used the same hyperparameters for all data sets. The latter affected the SVR performance throughout the case study, as the hyperparameter were not optimal in all cases. Also, in the case of stacked SVR, the combination of Gaussian kernels with a linear model led to a smoother trend than the one obtained with ERBFN, this affected the performance of the alternative ML technique in this case study. Nonetheless, the model built with the trend from ERBFN and SVR enhance the performance obtaining less misclassified material when compared to the other methodologies.
.

Conclusions

The present paper proposed to evaluate the applicability of ML in grade control models. This has been successfully achieved in the case study presented. Even though, the case study is performed over 10 different data sets, the hyperparameters for ML training are not set problem specific to each dataset considered. The performance of the prediction using a trend model obtained with ML, ERBFN algorithm,

outperformed the other methodologies compared. In this context, the grade estimation stage led to better predictions of the material extracted destinations on every area which are converted to revenue to the company. These results are encouraging relative to the applicability of ML in grade control context. Nonetheless, further analysis related to hyperparameter setting, dataset complexity and even other ML algorithms are necessary to fully understand and enhance modeling.

References

Aggarwal, C. C. (2018). NNs and Deep Learning. Springer Internetional Publishing. Ebook. https://doi.org/10.1007/978-3-319-94463-0.

Almeida, A. S. (1993). Joint Simulation of multiple variables with a Markov-type coregionalization model. Tese de Doutorado . Universidade de Stanford.

Almeida, A. S., Journel A. G. (1994). Joint simulation of multiple variables with a Markov-type coregionalization model. Mathematical Geology. 26. https://doi.org/10.1007/BF02089242

Chandra, A., Yao, X. (2006). Ensemble learning using multi-objective evolutionary algorithm. Journal of mathematical modelling and algorithms 5. https://doi.org/10.1007/s10852-005-0902-3

Dai, F., Zhou, Q., Lv, Z., Wang, X. and Liu, G. (2014) Spatial prediction of soil organic matter content integrating artificial NN and OK in Tibetan Plateau. Ecological Indicators 45. https://doi.org/10.1016/j.ecolind.2014.04.003.

Deutsch, C.V., Journel, A. G. (1998). GSLIB: a geostatistical software library and user's guide. New York (NY): Oxford University Press.

Dowd, P. A. and Saraç, C. (1994). A NN Approach to Geostatistical Simulation. Mathematical Geology, 26(4).

Gangappa, M., Mai, C. K., Sammulal, P. (2017) Techniques for machine learning based spatial data analysis: research directions. International Journal of Computer Applications, 170(1). https://doi.org/10.5120/ijca2017914643

Isaaks, E. H. (1990) The application of Monte Carlo methods to the analysis of spatially correlated data. PhD thesis. Stanford University, Unites States of America.

Journel, A. G., Huijbregts, C. J., (1978). Mining Geostatistics. Academic Press, London, England.

Journel, A. (1999). Markov Model for Cross-Covariances. *Mathmatical geology* , 955-964.

Kapageridis, I. K. (1999). Application of NNs systems to grade estimation from exploration data. (PhD). University of Nottingham.

Ma Z., Wang P., Gao Z., Wang R., Khalighi K. (2018) Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. PLoS One. 13(10). doi:10.1371/journal.pone.0205872

Maniar, H., Srikanth, R., Kulkarmi, M.S., Schlumberger, A. A. (2018) Machine Learning methods in geoscience. Society of Exploration Geophysicists. International Exposition and 88th Annual Meeting. https://doi.org/10.1190/segam/2018-2997218.1

Matheron, G. (1963) Principles of geostatistics. Economic Geology. 58(8). https://doi.org/10.2113/gsecongeo.58.8.1246

McCulloch, W.S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5(4). https://doi.org/10.1007/BF02478259.

Pedregosa, F., Voroquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchenary, E. (2011). Scikit-learn: ML in Python. Journal of ML. 12.

Samson, M. and Deutsch, C. (2018). Estimation with ML. CCG annual report 20. Paper 118.

Samson, M. and Deutsch, C. (2019). Elliptical Radial Basis Function vs. Radial Basis Functions. CCG annual report 21. Paper 110.

Samson, M. (2019). Mineral Resource Estimates with ML and Geostatistics. Master Thesis. Edmonton, Canada.

Tahmasebi, P. and Hezarkhani, A. (2012). A fast and independent architecture of artificial NN for permeability prediction. Journal of Petroleum Science and Engineering 86. doi: http://doi.org/10.1016/j.petrol.2012.03.019

Tomislav, H., Nussbaum, M., Wright, M. N., Heuvellink, G. B. M., Graler, B. (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables, PeerJ 6: e5518. https://doi.org/10.7717/perrj.5518

Vapnik, V. N. (1995) The nature of statistical learning theory. New York. Springer.

Vapnik, V.N. (1998) Statistical Learning Theory. New York. Wiley

Vasylchuk, Y., V. and Deutsch, C. V. (2017). IGC– Expected Profit. CCG Annual report 19. Paper 116.

Vasylchuk, Y. V. and Deutsch, C. V. (2018). IGCsystem overview and latest improvements. CCG annual report 20. Paper 305.

Vasylchuk, Y. V. and Deutsch, C. V., (2018). Optimization of Surface Mining Dig Limits with Realistic Selectivity. CCG annual report 20. paper 307.

Vasylchuk, Y. V. (2019). Advanced Grade Control with Multivariate Geostatistics, Blast Movement Modeling and Optimized Dig Limits. PhD Thesis. Edmonton, Canada.

Verly, G. (2005). Grade control classification of ore and waste: a critical review of estimation and simulation-based procedures. Mathematical Geology 37. doi: http://doi.org/10.1007/s11004-005-6660-9.

Walch, A. Castello, R., Mohajeri, N., Guinard, F., Kanevski, M. Scartezzini, JL., (2019) Spatio-temporal modelling and uncertainty estimation of hourly global solar irradiance using extreme learning machines. Energy Procedia. 158. https://doi.org/10.1016/j.egypro.2019.01.219

Witten, I. H., Frank, E., Hall, M.A., Pal, C.J., (2017) Ensemble Learning. In Data Mining 4th edition. Springer. doi: hhttps://doi.org. 10.106/B978-0-12-80249-5.0012-X

Wolpert, D. H., (1992). Stacked generalization. Neural Networks (5)2. Doi. https://doi.org./10.106.50803-6080(05)80023-1

Xu, W., Tran, T. T., Srivastava, R. M., & Journel, A. G. (1992). Integrating Seismic Data in Reservoir Modelling: The Collocated Cokriging Alternative. Society of Petroleum Engineers, pp. 833-842