mathematics

MDPI

Article

Comparison of Trivariate Copula-Based Conditional Quantile Regression Versus Machine Learning Methods for Estimating Copper Recovery

Heber Hernández [1], Martín Alberto Díaz-Viera [2], Elisabete Alberdi [3] and Aitor Goti [4,*]

[1] Facultad de Ingeniería, Universidad Santo Tomás, Ejército Libertador 146, Santiago 8370003, Chile; hhernandez9@santotomas.cl
[2] Instituto Mexicano del Petróleo, Eje Central Lázaro Cárdenas No. 152, Ciudad de Mexico 07730, Mexico; mdiazv@imp.mx
[3] Department of Applied Mathematics, University of the Basque Country UPV/EHU, 48013 Bilbao, Spain; elisabete.alberdi@ehu.eus
[4] Department of Mechanics, Design and Organization, University of Deusto, 48007 Bilbao, Spain
* Correspondence: aitor.goti@deusto.es

Abstract: In this study, an innovative methodology using trivariate copula-based conditional quantile regression (CBQR) is proposed for estimating copper recovery. This approach is compared with six supervised machine learning regression methods, namely, Decision Tree, Extra Tree, Support Vector Regression (linear and epsilon), Multilayer Perceptron, and Random Forest. For comparison purposes, an open access database representative of a porphyry copper deposit is used. The database contains geochemical information on minerals, mineral zoning data, and metallurgical test results related to copper recovery by flotation. To simulate a high undersampling scenario, only 5% of the copper recovery information was used for training and validation, while the remaining 95% was used for prediction, applying in all these stages error metrics, such as R^2, MaxRE, MAE, MSE, MedAE, and MAPE. The results demonstrate that trivariate CBQR outperforms machine learning methods in accuracy and flexibility, offering a robust alternative solution to model complex relationships between variables under limited data conditions. This approach not only avoids the need for intensive tuning of multiple hyperparameters, but also effectively addresses estimation challenges in scenarios where traditional methods are insufficient. Finally, the feasibility of applying this methodology to different data scales is evaluated, integrating the error associated with the change in scale as an inherent part of the estimation of conditioning variables in the geostatistical context.

Keywords: metallurgical copper recovery; copula; quantile regression; kernel smoothing; machine learning

MSC: 62H05

Academic Editors: Víctor Amor-Esteban, David Almorza-Gomar and Raymond Lee

Received: 29 December 2024
Revised: 28 January 2025
Accepted: 6 February 2025
Published: 10 February 2025

Citation: Hernández, H.; Díaz-Viera, M.A.; Alberdi, E.; Goti, A. Comparison of Trivariate Copula-Based Conditional Quantile Regression Versus Machine Learning Methods for Estimating Copper Recovery. *Mathematics* **2025**, *13*, 576. https://doi.org/10.3390/math13040576

1. Introduction

The design and planning of a mine is largely based on the block-discretized resource model [1–5]. The economic valuation of the blocks depends on geological, geometallurgical, market, environmental, and sociopolitical parameters [6]:

$$Block_{value} = (P - C_s) \cdot T_b \cdot R_b \cdot G_b - (T_b \cdot C_b), \quad (1)$$

where P is the price of the metal, which depends on the global market [7,8], C_s the cost of sale, T_b is the tonnage and depends on the size and density of the block, R_b is the estimated metallurgical recovery of the block, G_b is the mineral grade, traditionally estimated with geostatistical methods [9–11], and C_b the total extraction and processing costs of the block, which are positively related to the price [12].

In terms of complexity to estimate these parameters, the metallurgical responses, which are broadly defined in [6,13–15], are the ones at the top of the list. This is because samples with metallurgical response information are scarce, expensive to obtain, highly undersampled, and have nonlinear relationships with primary geological variables, in addition to their support generally being uneven [16–26].

In porphyry copper deposits, characterized by disseminated and veinlet mineralization associated with igneous intrusions, copper metallurgical recovery (CR) depends on various intrinsic rock characteristics, such as mineralogical composition, the presence of contaminants, mineral grades, textures, and grain size, among others [13,27,28]. These properties characterize the geometallurgical domains, defined as three-dimensional regions within the deposit where metallurgical responses are expected to present a homogeneous distribution [29,30]. In the particular case of copper sulfides, CR is traditionally obtained by flotation [31,32], which is expressed as $R^{C}_{CR}u$ in Equation (2), and is defined as the relationship between the mass of metal in the concentrate m_c^{Cu} and the mass of metal in the feed m^{Cu} [33]. This variable is nonadditive, so the use of traditional geostatistical techniques for its estimation causes biases in the mixture [34,35]:

$$R^{Cu}_{CR} = \frac{m_c^{Cu}}{m^{Cu}}. \tag{2}$$

Since it has nonlinear relationships with more densely sampled predictor variables, the use of geostatistical co-estimations and multiple linear regressions are also not appropriate.

Currently, the trend in geometallurgy is to use Machine Learning (ML) models, which combine primary geological variables and mine information, with metallurgical responses [36]. These models are divided into two types: unsupervised, which allow geometallurgical domains to be identified and defined through similarity analysis, and supervised, which use nonlinear regression algorithms to predict metallurgical responses once said domains are established [37]. The sequential combination of these two approaches is presented in workflows depending on the particular case to be treated [38].

For example, Ref. [39] compared hierarchical clustering, k-means, and self-organizing maps to define geometallurgical domains in an iron deposit in Iran. Similarly [27] applied k-means on three principal components to establish domains related to metallurgical responses, which were subsequently estimated using a Random Forest model at the Paracatu mine, Brazil. In another study, Ref. [40] used a spatially corrected fuzzy clustering method to define geometallurgical domains in a 2D section of a synthetic case developed by [22], which is also used in this study, but extended to 3D. Ref. [41] applied the Gaussian Mixture Model to define geometallurgical domains in the Carajás iron ore deposit, Brazil. For their part, Ref. [42] used neural networks to estimate metallurgical response variables from mining operation data at the Tropicana mining complex, Australia. Likewise, Ref. [28] used the Extreme Gradient Boosting algorithm to predict the metallurgical recovery of copper in the Tizert deposit, Morocco, highlighting that the most relevant predictive characteristics were the geological unit, the copper-bearing mineral, and the oxidation rate. As demonstrated by these cases, the use of ML in geometallurgy has established itself as an acceptable and constantly growing solution.

In this paper, we present an innovative methodology for CR estimation using a dependency model based on multivariate copulas [43–46] and conditional quantile

regression. Multivariate copulas are used to model dependency relationships between random variables, independently of the behavior of their marginal distributions. Copulas capture complex relationships, including nonlinear dependencies, which are common in geometallurgical data. On the other hand, conditional quantile regression is a statistical method that allows estimating a variable for a given quantile conditionally using the rest of the variables of the joint probability distribution function, which is here constructed using a copula model that couples the marginal distributions. This method is particularly useful for analyzing and modeling phenomena under extreme or heterogeneous conditions.

This methodology, previously applied in a 2D case with a bivariate copula model [23], is extended here to a 3D case using a trivariate copula. The proposed methodology is compared with six supervised ML regression methods for CR estimation using mostly sampled predictor variables. This comparison is particularly relevant in the context of CR in porphyry copper deposits, where the variability of metallurgical responses is influenced by a wide range of geological factors. ML methods, although effective in many cases, face challenges due to the high nonlinearity and scarcity of CR-specific data. In this regard, the proposed model based on trivariate copulas offers an interesting alternative by explicitly incorporating the dependence between predictor variables and the variable of interest. Moreover, its ability to handle dependency structures and spatial heterogeneity in geometallurgical data makes it a useful tool for improving CR estimation in situations where traditional models and conventional ML methods may not be as effective.

The structure of the paper is as follows. In the Section 2, the general proposed methodology is described. The Sections 3 and 4 present the multivariate copula-based conditional quantile regression and the supervised machine learning regression methods, respectively. The Section 5 shows the results of the application of both approaches to the case study. In the Section 6, the comparison of the performance of the two methods is discussed (see also in Section 7), and finally, in the Section 8, the conclusions are given.

2. Methodology

The methodology is organized in several stages, as shown in Figure 1. First, data collection, exploratory analysis, and variable selection that define the geometallurgical model are carried out. Next, the selected variables are standardized, both for the copula-based dependency model and for the training of the ML methods. Once both approaches have been validated, the CR is estimated in places where measurements are not available, conditioned on the predictor variables. Finally, the results obtained from both approaches are compared. The general approach of this methodology reflects a trend in geometallurgical workflows, as observed in the case of the Olympic Dam deposit in Australia. There, through detailed mineralogical and geochemical characterization of both ore and waste, metallurgical plant responses are estimated to optimize its performance [47].

- Exploratory Data Analysis (EDA): This is the first stage of the workflow after data collection. Its primary goal is to identify patterns, trends, and outliers through the use of descriptive statistics and visual tools. A fundamental part of this analysis is the study of multivariate dependence, which is based on the calculation of the rank correlation Spearman coefficient ($\rho_{s_{ij}}$). This coefficient is organized in a symmetric matrix of size $N \times N$, where each element (i, j) represents the correlation between the variables X_i, X_j. The diagonal elements $(i = j)$ have a value equal to 1, since a variable is perfectly correlated with itself. Equation (3) shows how $\rho_{s_{ij}}$ is calculated:

$$\rho_{s_{ij}} = \frac{Cov(R[X_i], R[X_j])}{\sigma_{R[X_i]}\sigma_{R[X_j]}}, \forall i, j \in 1, 2, ..., N, \tag{3}$$

where $Cov(R[X_i], R[X_j])$ is the covariance between the rank variables $R[X_i]$ and $R[X_j]$, while $\sigma_{R[X_i]}$ and $\sigma_{R[X_j]}$ are the standard deviations of the rank variables. This approach allows nonlinear interactions between different characteristics to be evaluated more effectively. The result of this stage is the identification of key relationships and the selection of features, which forms the basis for defining geometallurgical domains.

- Definition of geometallurgical domains: Based on the dependency analysis carried out on the characteristics of the sample and the selection of those that influence the CR, the geometallurgical domains are constructed. This process can be carried out through various methodologies proposed in studies such as [27,39,41,48].
- Data Preprocessing: The selected features undergo a transformation process to standardize their mean and variance. This step is essential when working with ML models, as it ensures optimal conditions for their operation [49,50]. Although it is not a mandatory requirement for copula models, it is recommended to carry out this normalization, since it simplifies the modeling of marginal distributions and contributes to greater consistency in the results. The transformation formula used is:

$$z_i = \frac{x_i - \mu}{\sigma}, \tag{4}$$

where x_i is the value of the untransformed variable, μ and σ are the mean and standard deviation, respectively, and z_i the transformed value of the variable.

- Split Data: Typically, for ML methods, the sample is partitioned into two data subsets: one for training and the other for testing, using a proportion of 70/30%.
- Modeling:
  - Dependency Modeling: The joint probability distribution function is modeled by a multivariate copula and its marginals using a nonparametric approach applying the kernel smoothing method.
  - Model Training: ML methods are trained with different hyperparameters until acceptable performance metrics are obtained for the training sample data.
- Validation: At this stage, the previously obtained models are validated using the performance metrics for the testing sample data.
- Estimation: ML and CBQR methods are applied to predict copper recovery, in 95% of the data samples, that are not used in the previous modeling and validation stages.
- Comparison: The comparison of the results obtained in the previous stage is carried out in terms of the performance metrics.

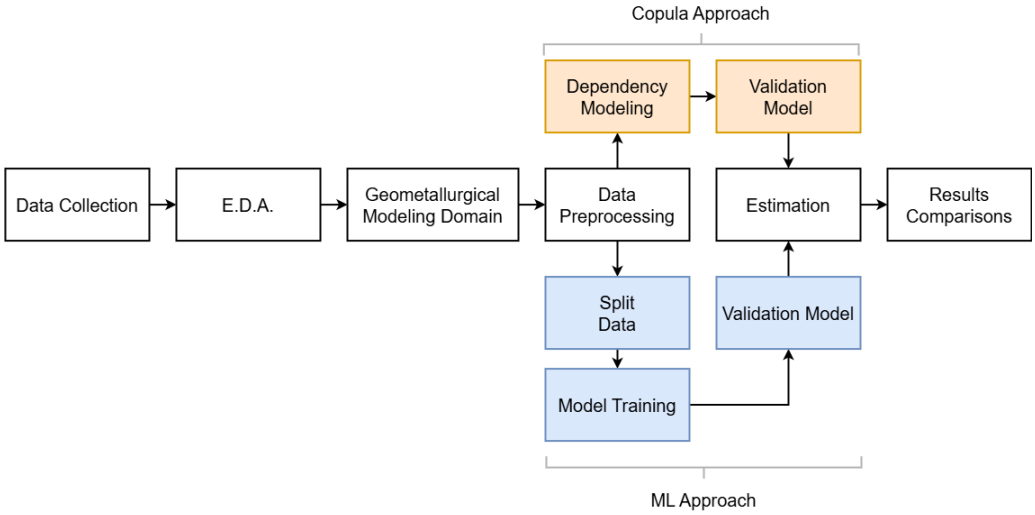

Figure 1. Methodological workflow.

3. Copula-Based Conditional Quantile Regression

Copulas are statistical functions that facilitate the modeling of complex nonlinear dependencies between multiple variables by connecting their univariate cumulative distribution functions [44,45,51–55].

Mathematically, for a random vector $(X_1, X_2, \ldots, X_d)$ with a joint distribution function $H(x_1, x_2, \ldots, x_d)$ and marginal distributions $F_1(x_1), F_2(x_2), \ldots, F_d(x_d)$, Sklar's theorem [54], refined by Nelsen in 2006 [55], asserts that a copula $C : [0,1]^d \to [0,1]$ exists such that:

$$H(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)), \tag{5}$$

where C encapsulates the dependence structure between variables, while $F_i(x_i)$ represent the marginal distributions. When the F_i values are continuous, the copula C is unique.

3.1. Key Properties of Copulas

The fundamental properties of copulas include:

- **Marginality:** $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$, for any $u_i \in [0,1]$.
- **Boundary Conditions:** $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$, for any $u_i \in [0,1]$.
- **Nonnegative Volume:** The function C must be d-increasing, meaning the volume it describes within any hyperrectangle in $[0,1]^d$ is nonnegative.

To ensure the nonnegativity of this volume, an auxiliary function $h : [0,1]^d \to \mathbb{R}$ is introduced. This function is d-increasing if, for any hyperrectangle $[a, b]$, the following condition holds:

$$V_h([a,b]) = \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} h(t)\, dt \geq 0. \tag{6}$$

where $h(t)$ reflects the boundary behavior of the margins based on the variables $t_1, \ldots, t_d$.

3.2. Multivariate Copula-Based Conditional Quantile Regression

Quantile regression aims to model conditional quantiles of a response variable, offering a comprehensive analysis of potential outcomes given predictor variables. Multivariate Conditional Quantile Regression (CBQR) extends classical quantile regression by utilizing copulas to handle multivariate outcomes.

The objective of multivariate CBQR is to predict the quantile of a response variable Y conditioned on predictor variables $X_1, \ldots, X_d$, where $Y \sim F_Y$ and $X_j \sim F_j$, $j = 1, \ldots, d$. The conditional quantile function for $\alpha \in (0,1)$ is given as:

$$q_\alpha(X_1, \ldots, X_d) = F^{-1}_{Y|X_1,\ldots,X_d}(\alpha|X_1, \ldots, X_d). \tag{7}$$

Using the probability integral transforms $V = F_Y(Y)$ and $U_j = F_j(X_j)$, the conditional distribution is expressed as:

$$F_{Y|X_1,\ldots,X_d}(y|X_1, \ldots, X_d) = C_{V|U_1,\ldots,U_d}(V|U_1, \ldots, U_d). \tag{8}$$

Then, the inverse function gives:

$$F^{-1}_{Y|X_1,\ldots,X_d}(\alpha|X_1, \ldots, X_d) = F_Y^{-1}\left(C^{-1}_{V|U_1,\ldots,U_d}(\alpha|U_1, \ldots, U_d)\right). \tag{9}$$

Estimates for the conditional quantile function can be derived by estimating F_Y, F_j, and the copula C, resulting in:

$$\hat{q}_\alpha(X_1, \ldots, X_d) = \hat{F}_Y^{-1}\left(\hat{C}^{-1}_{V|U_1,\ldots,U_d}(\alpha|\hat{U}_1, \ldots, \hat{U}_d)\right), \tag{10}$$

where $\hat{U}_j = \hat{F}_j(x_j)$ represents the estimated probability integral transforms of X_j.

3.3. Kernel Smoothing in Copula Estimation

Kernel smoothing is a nonparametric technique that estimates unknown density functions by averaging data points over a predefined "bandwidth" using a kernel function.

For multivariate data, the kernel density estimate (KDE) is given by:

$$\hat{f}(x_1, \ldots, x_d) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \frac{1}{h_j} K\left(\frac{x_j - X_{ij}}{h_j} \right), \tag{11}$$

where K is the kernel function, h_j are the bandwidths for each dimension, and X_{ij} is the j-th component of the i-th data point.

The copula density is estimated from the joint and marginal densities as:

$$\hat{c}(u_1, \ldots, u_d) = \frac{\hat{f}(\hat{F}_1^{-1}(u_1), \ldots, \hat{F}_d^{-1}(u_d))}{\prod_{j=1}^{d} \hat{f}_j(\hat{F}_j^{-1}(u_j))}. \tag{12}$$

The copula function is then derived by integrating the estimated copula density using numerical techniques such as the trapezoidal or Simpson's rule. To simulate data from the copula:

1. Generate samples from a uniform distribution.
2. Transform these samples using the inverse of the copula function.

The implementation was conducted using Python and the `OpenTURNS` library [56].

4. Machine Learning Methods

Machine learning methods for regression are designed to predict a continuous outcome variable based on one or more predictor variables. Six supervised machine learning regression methods, i.e., Decision Tree (DT), Extra Tree (ET), Support Vector Regression (SVR), both linear (SVRL) and with epsilon (SVRE), Multilayer Perceptron (MLP), and Random Forest (RF), are applied in this work. These methods model the relationship between the predictors and the target variable by minimizing the prediction error. Below is a brief description of these machine learning methods for regression.

4.1. Decision Tree (DT)

Decision Trees are nonlinear models that partition the data into subsets recursively based on a selected feature that achieves the largest reduction in target variance [57]. At each node, the data are split based on a feature threshold, aiming to create homogeneous subsets in terms of the target variable:

- Prediction at each leaf node is often the average of the target values of samples within the node.
- Prone to overfitting, especially if the tree is very deep.
- Easily interpretable but sensitive to changes in the data.

4.2. Support Vector Regression (SVR)

Support Vector Regression extends SVM concepts to regression problems [58]:

- Aims to fit the error within a certain threshold and focuses on minimizing the model coefficients to enhance generalization.
- Employs kernel functions (e.g., linear, polynomial, and RBF) to manage nonlinear relationships by transforming data into a higher-dimensional space.
- Robust against overfitting in high-dimensional spaces.

4.3. Multilayer Perceptron (MLP)

Multilayer Perceptrons are types of neural networks that include [59]:

- An input layer, one or more hidden layers, and an output layer.
- Utilization of backpropagation for training, adjusting weights to minimize prediction errors.
- Activation functions (e.g., ReLU, sigmoid) introduce nonlinearity, enabling the network to learn complex patterns.
- Effective for complex functions but they are computationally expensive and require careful parameter tuning.

4.4. Random Forest (RF)

Random Forests are ensemble learning methods that improve on decision trees [60]:

- Builds multiple decision trees and merges them to enhance accuracy and stability.
- Takes the average of outputs from all trees for final prediction in regression tasks.
- Offers better resistance to overfitting than individual decision trees.
- Effective for a wide range of problems but can be less interpretable due to its ensemble nature.

The computational implementation was carried out in a notebook using the open source library `scikit-learn` [61] with the `Python` programming language.

4.5. Error Performance Metrics Definitions

- Determination coefficient:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{13}$$

where y_i are the actual values, $\hat{y}_i$ are the predictions, and $\bar{y}$ is the average of the actual values.

- Maximum relative error:

$$MaxRe = \max_i \left(\frac{|y_i - \hat{y}_i|}{|y_i|} \right), \tag{14}$$

- Mean absolute error:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \tag{15}$$

- Mean squared error:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{16}$$

- Median absolute error:

$$MedAE = median|y_i - \hat{y}_i|, \tag{17}$$

- Mean absolute percentage error:

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \tag{18}$$

where $y_i \neq 0$.

5. Case Study

For this application, a freely accessible synthetic dataset is used, specifically designed for academic purposes and oriented towards geometallurgical research. It provides a detailed representation of a porphyry copper type mineral deposit. The methodology for constructing this dataset can be found in [22]. The dataset consists of 4597 records from mining exploration drillholes, each containing geochemical data for minerals such as clays, chalcocite, bornite, chalcopyrite, tennantite, molybdenite, and pyrite. Additionally, it includes information on copper grade, molybdenum grade, geometallurgical units (GMU), and the results of metallurgical tests for copper recovery by flotation. This dataset is suitable for modeling CR from mineral concentrations, as suggested by [62], which validates its suitability for the research presented.

5.1. Geometallurgical Units Definition

In relation to the geometallurgical units (GMU) already defined in this database, GMU-1 corresponds to oxidized copper ores, located at the water table, with evidence of leaching that reflects supergenic alteration processes (see Figure 2).

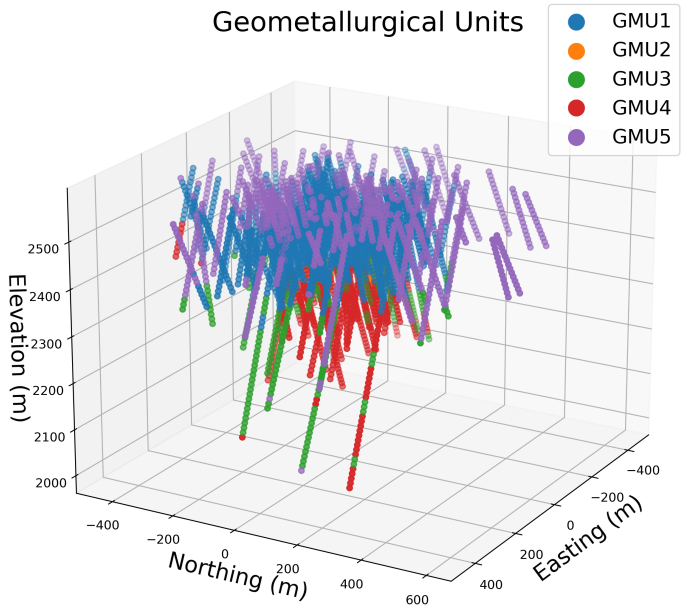

Figure 2. Three-dimensional visualization GMUs.

In the processing of copper ores, oxidized ores are commonly soluble in acidic solutions and the metal is recovered through heap leaching. Below this is the GMU-2, dominated by secondary sulfides such as chalcocite, forming a layer of enrichment sulfides generated by the remobilization and precipitation of copper. In depth, GMU-3 and GMU-4 represent primary hypogenic sulfides, differentiated by the proportion of chalcopyrite and pyrite: GMU-3 has a higher concentration of these minerals, while GMU-4 has a lower proportion. Sulfide minerals do not dissolve easily in acidic solutions like copper oxides, and must be treated through flotation. Precisely these last two units are the ones that present the greatest recovery of copper by flotation. Finally, GMU-5 corresponds to waste and gravel without economic copper content, constituting the superficial sterile material of the deposit.

In order to estimate the CR, its relationship with the minerals and elements present in the sample is analyzed. Although the geometallurgical units provide a framework for characterizing the deposit, the data used in building predictive models are not limited exclusively to these units independently. This is supported by the marked statistical

relationship that exists between minerals and CR in all GMUs (see Figure 3). Furthermore, that is why in this application case the models are built using all the available information.

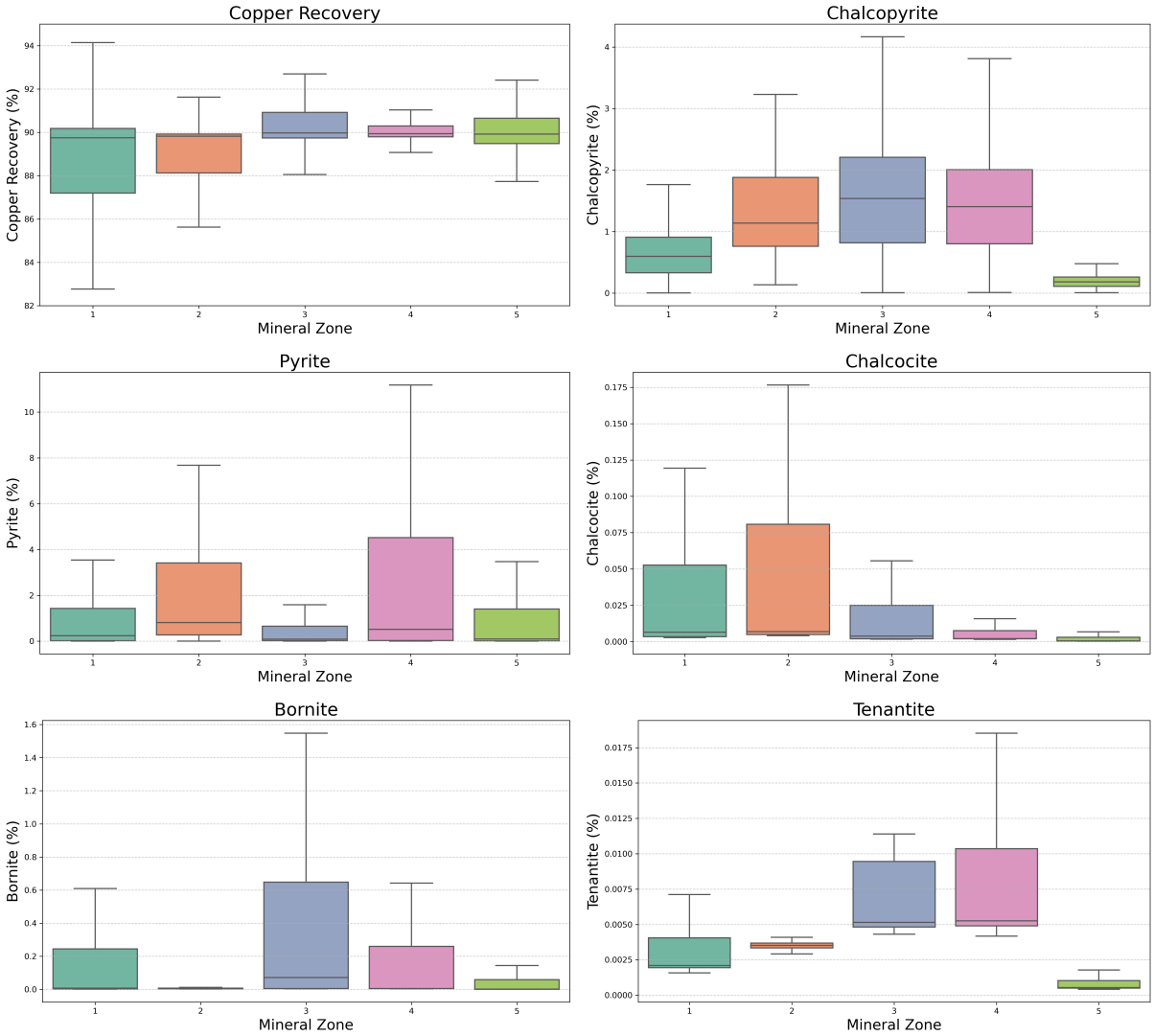

Figure 3. Boxplot for minerals by GMU.

To directly evaluate the results of modeling and estimation, a 5% subsampling was applied to the available CR data at the drillhole samples. This approach aims to replicate a realistic scenario in which metallurgical response is significantly underrepresented compared to geological information. A similar case is described by [33], who worked with a real dataset where only 3% of the information corresponded to metallurgical recovery data. This situation enables testing the robustness of the copula based dependency model under conditions of limited information. Subsequently, the results of the median regression conditional on minerals will be compared with the 95% of the actual CR data reserved specifically for this exercise.

Figure 4 shows the distribution of a 5% CR sample and the estimation locations in a sulfide geometallurgical unit.

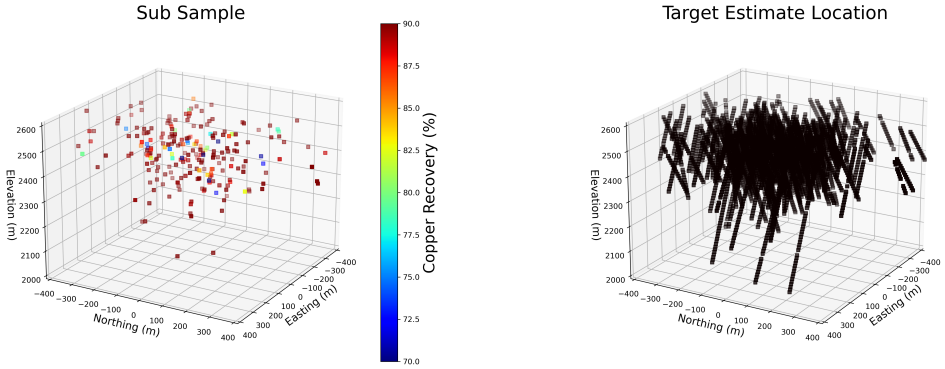

Figure 4. 5% CR sample distribution (**left**) and location target for estimation (**right**).

5.2. Exploratory Data Analysis

A basic statistics summary for the 5% CR sample, 100% CR sample, and 5% CR nonconditional simulation are given in Table 1. Figure 5 shows the consistency of the 5% and 100% histograms for the CR of chalcocite and bornite, respectively.

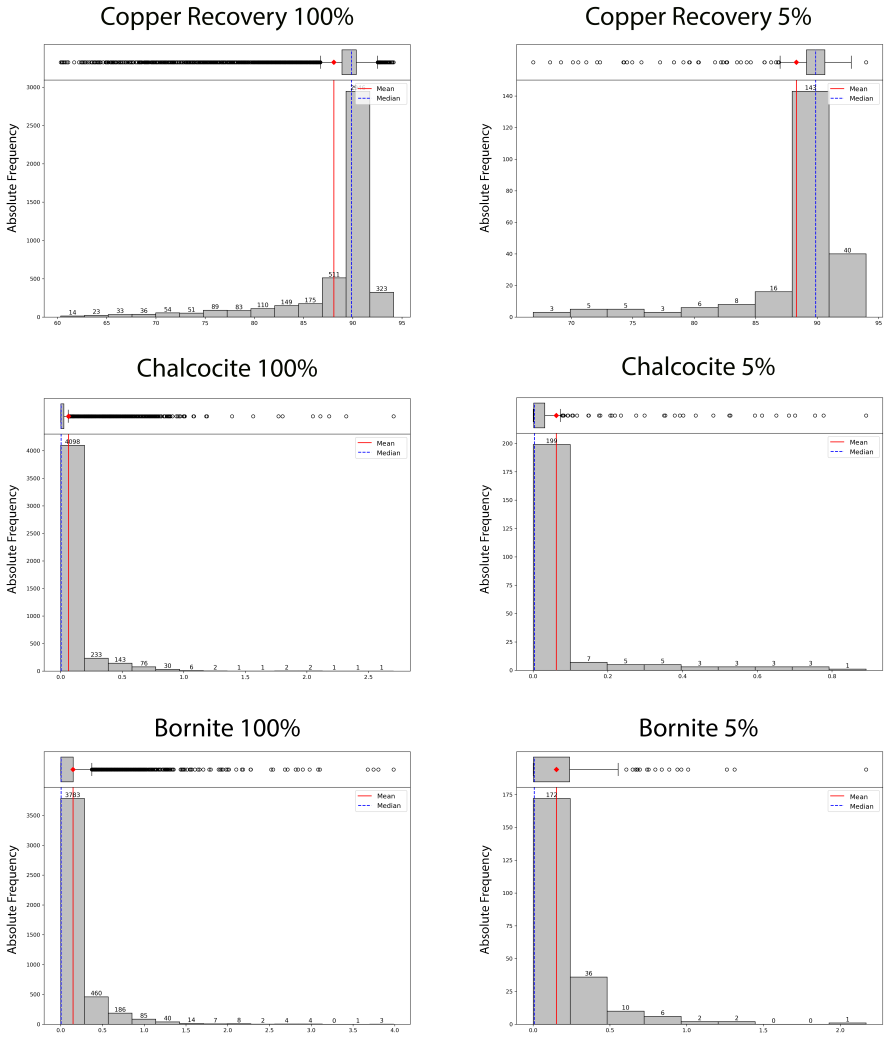

Figure 5. Histograms and boxplots for copper recovery, chalcocite and bornite from 100% and 5% samples, respectively.

Table 1. Statistics summary for the 5% CR sample, 100% CR sample, and 5% CR nonconditional simulation.

Statistics	5% CR Sample	100% CR Sample	5% CR Simulated
Count	229	4.597	229
Mean	88.31	88.08	88.25
Std	4.87	5.30	4.61
Min	66.90	60.34	67.23
Q1	89.13	88.91	88.31
Q2	89.89	89.87	89.89
Q3	90.62	90.35	90.49
Max	93.98	94.15	94.32
Kurt	6.26	6.26	8.89
Skew	−2.59	−2.59	−2.46

Variable Selection

To build the trivariate copula model, those primary variables are selected that have a high dependence on the CR and, at the same time, are minimally interdependent on each other to avoid redundancies in the model. In this process, the Spearman correlation coefficient was used, suitable for capturing nonlinear relationships between variables.

The greatest dependencies were observed in bornite and chalcocite, with correlations of 0.49 and −0.59, respectively, in relation to CR. Furthermore, these variables are not redundant with each other, since they have a low mutual correlation of 0.18 (see Figures 6 and 7). The most informative variables are chalcocite and bornite (see Figure 8).

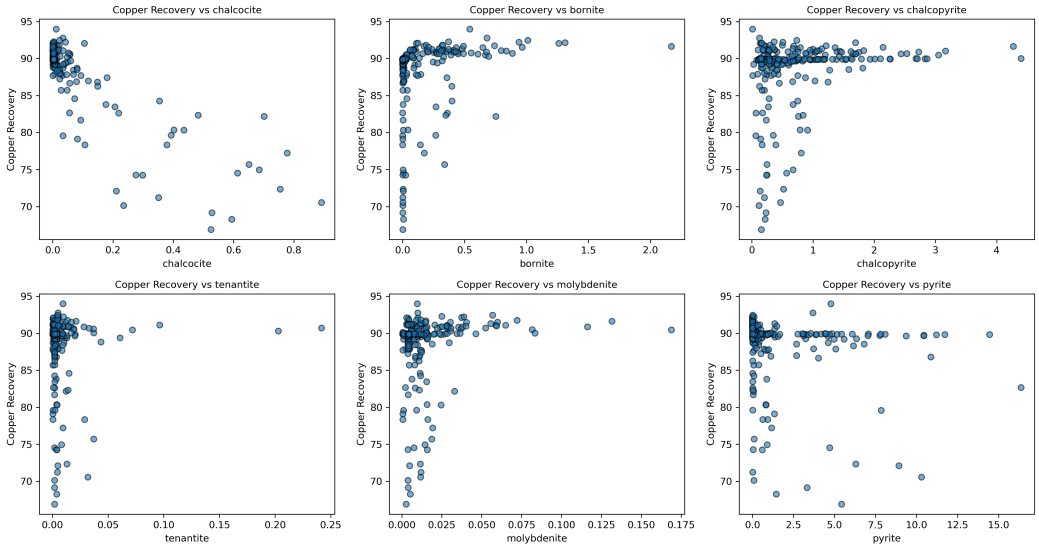

Figure 6. Scatter plot of CR vs. geochemical minerals.

5.3. Modeling

5.3.1. Dependency Modeling

Marginal distributions are modeled using the nonparametric smoothed kernel method, which allows estimating both the probability density function (PDF) and the cumulative distribution function (CDF) directly from the data, without assuming a specific theoretical form. To do this, a Student function is used, which defines a window around each data point and smoothes the PDF estimate.

The modeling is carried out on a transform of the original variables, which, although not strictly necessary, improves the fit in cases of high skewness and kurtosis.

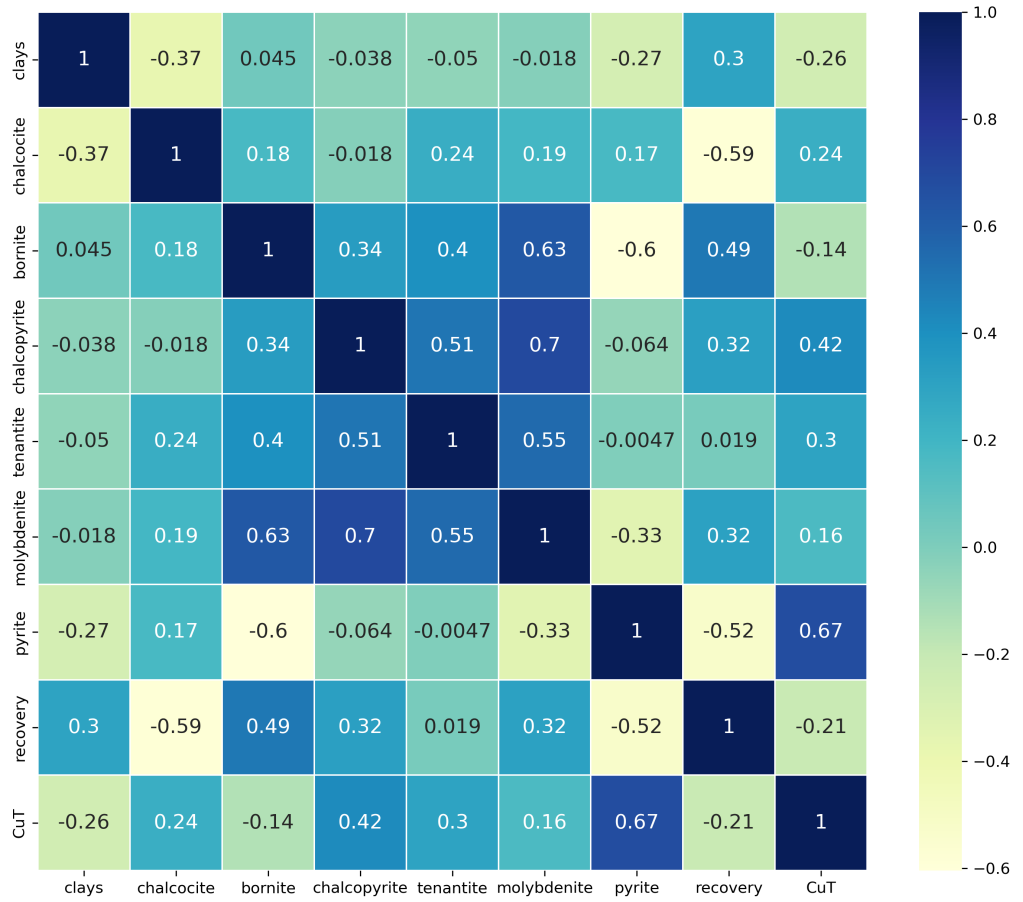

Figure 7. Spearman correlation matrix.

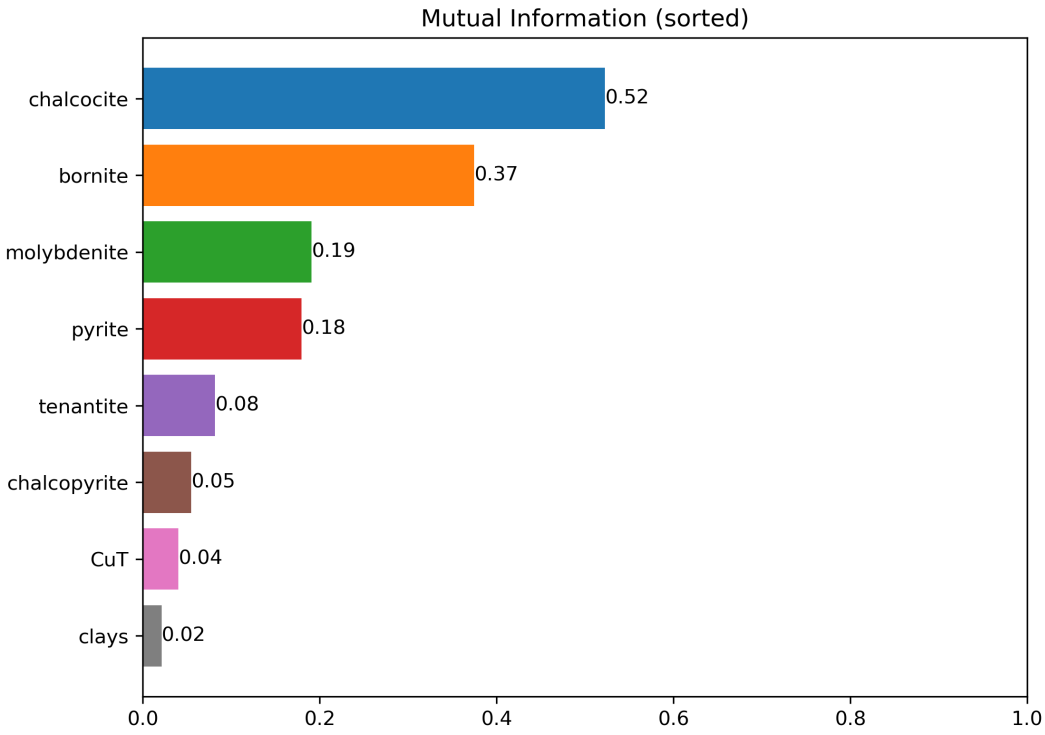

Figure 8. Mutual information for variable selection.

The construction of the trivariate copula is carried out by combining the fitted marginal distributions (Figure 9) and a dependence model defined by the copula (Figure 10), where the Epanechnikov kernel is used, which demonstrated better performance compared

to other tested functions, such as Laplace, Student, Chi-square, and Exponential. This approach ensures that the final model respects both the individual marginal properties and the joint dependence structure.

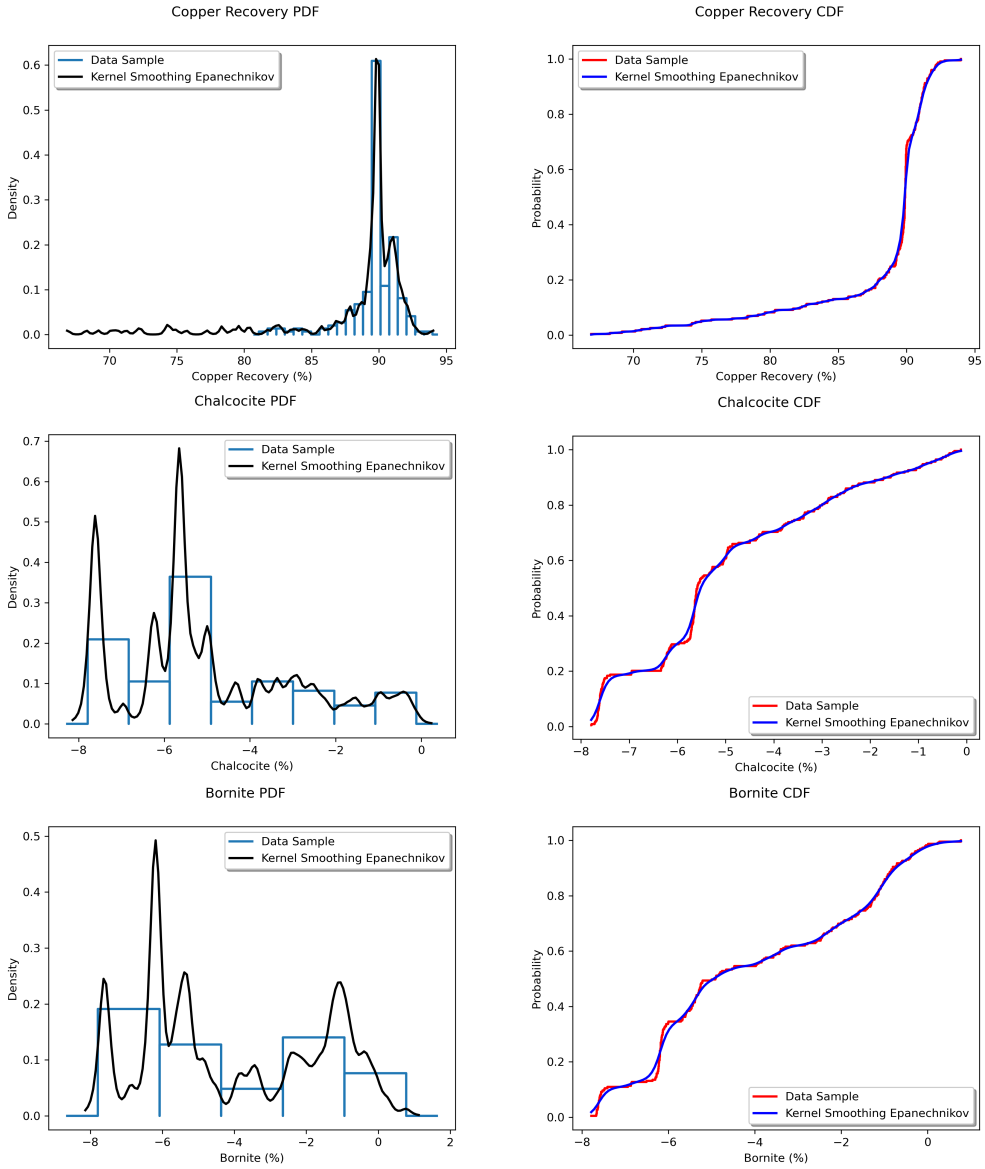

Figure 9. Nonparametric fitting of marginals using the Kernel Smoothing method.

5.3.2. Model Training

In the DT, ET, and RF models, the *max depth* hyperparameter defines the maximum depth of the decision trees, thereby limiting the model's complexity to reduce the risk of overfitting. In the case of RF, additional hyperparameters were used, including *n estimators*, which specifies the number of trees in the forest, and *bootstrap*, indicating that data are resampled with replacement during model construction, which enhances the algorithm's robustness and generalization ability.

For the SVRL model, the hyperparameters *max iter* were set to define the maximum number of iterations in the optimization process, and *tol*, which determines the tolerance level for convergence, impacting the model's fitting precision. In the case of SVRE, in addition to *tol*, the following hyperparameters were included: *kernel*, configured as RBF (Radial Basis Function), which defines the kernel function used to transform the data into a higher dimensional space; gamma, set to scale, controlling the influence of each data

point on the model; and *C*, which balances model regularization by establishing a trade-off between minimizing training error and maximizing the margin between observations.

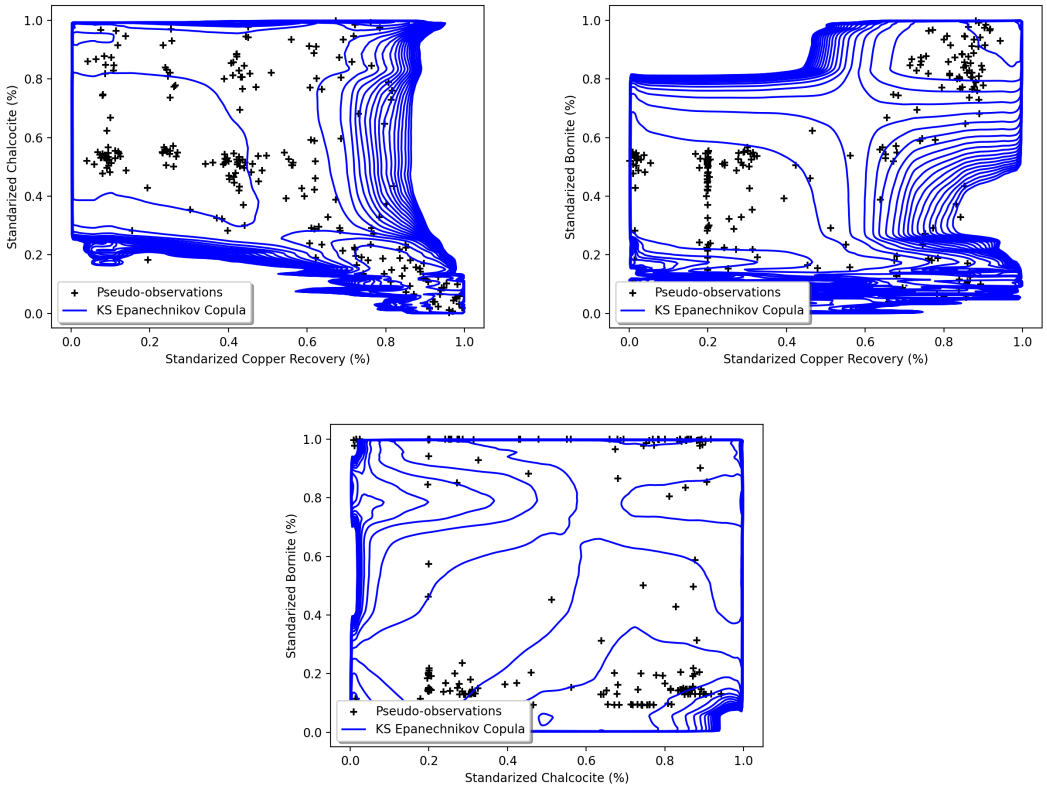

Figure 10. Nonparametric fitting of marginals using the kernel smoothing method.

Regarding the MLP model, the neural network configuration included the *hidden layer* hyperparameter, which specifies the number of hidden layers with 200 and 100 neurons, respectively, directly affecting the model's representation capacity. The *max iter* hyperparameter limits the maximum number of iterations during training, while *solver*, set to adam, specifies the optimization algorithm used. Additionally, the *learning rate* hyperparameter was set to 0.001, establishing the initial learning rate, and *alpha*, representing a regularization term to prevent overfitting. Lastly, *batch size* was set to auto, allowing the model to automatically determine the batch size for each iteration, and *activation* was defined as ReLU, selecting the Rectified Linear Unit activation function, which facilitates the representation of nonlinear relationships. Table 2 shows a summary of the hyperparameters used in the machine learning methods.

Table 2. Summary of the hyperparameters used in the machine learning methods.

ML	Hyperparameters	Value
DT	max depth	7
ET	max depth	7
SVRL	max iter	1000
	tol	0.001
SVRE	kernel	rbf
	gamma	scale
	tol	0.001
	C	80

Table 2. *Cont.*

ML	Hyperparameters	Value
DT MLP	hidden layer	2
	n neurons	(200, 100)
	max iter	5000
	solver	adam
	learning rate	0.001
	alpha	0.0001
	batch size	auto
	activation	relu
RF	n estimators	1000
	max depth	7
	bootstrap	True

The perfomance metrics for CBQR and ML methods during the training stage are given in Table 3.

Table 3. Perfomance metrics for CBQR and ML methods during the training stage.

	CBQR	DT	ET	SVRL	SVRE	MLP	RF
Train. score (%)	1.00	1.00	0.98	0.64	0.89	0.95	0.98
Train. Time (s)	0.002	0.005	0.002	0.006	0.008	3.634	0.740
R^2	0.89	0.93	0.79	0.68	0.87	0.88	0.93
MaxRE	0.11	3.85	9.13	11.35	7.22	7.12	3.95
MAE	0.78	0.72	1.12	1.30	0.83	0.82	0.75
MSE	2.05	1.29	4.18	6.33	2.57	2.24	1.43
MedAE	0.38	0.38	0.36	0.50	0.30	0.34	0.40
MAPE	0.01	0.01	0.01	0.02	0.01	0.01	0.01

5.4. Validation

Through a nonconditional simulation based on the trivariate copula model, and the associated marginal distributions, synthetic data are generated following the joint dependence structure. This procedure allows obtaining samples that respect the original statistical properties of the modeled system. The process develops as follows:

- Generation of Uniform Samples: Initially, random values uniformly distributed in the interval $[0, 1]$ are generated for each dimension of the trivariate space. These values represent the cumulative probabilities (CDF) corresponding to the marginal variables.
- Modeling of Dependence through Copula: The copula transforms these uniform samples to ensure that they respect the joint dependency structure defined in the model. This step ensures that relationships between variables, such as correlations or nonlinear dependence patterns, are reflected in the simulated data.
- Transformation to the Marginals: The simulated values are converted from the cumulative probabilities to the original scales of the marginal variables using the inverse cumulative distribution functions (ICDF) of each marginal. This reconstructs the individual characteristics of each variable while preserving their joint relationship. Finally, the simulation generates a synthetic dataset that can be compared with the observed data to validate the model. This process involves analyzing whether the descriptive statistics and the dependence between variables in the simulated sample are consistent with those of the original data (see Table 3 and Figure 11).

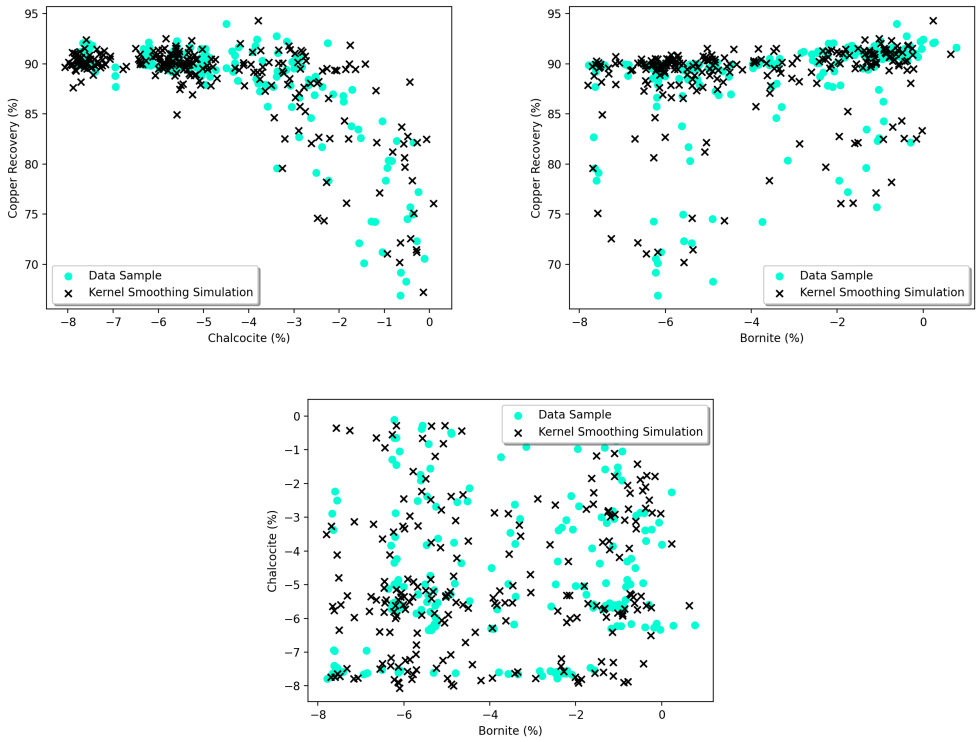

Figure 11. Simulation from the copula model versus 5% CR sample.

5.5. Estimation

Once the copula is modeled with kernel smoothing, the CBQR method conditioned with the variables bornite and chalcocite is applied to estimate the median CR in unsampled localities. The perfomance metrics of both the CBQR and the ML methods are given in Table 4.

Table 4. Perfomance metrics for CBQR and ML methods during the prediction stage.

	CBQR	DT	ET	SVRL	SVRE	MLP	RF
R^2	0.77	0.62	0.54	0.49	0.71	0.71	0.67
MaxRE	22.66	26.39	25.13	65.18	24.35	23.51	25.42
MAE	1.11	1.45	1.97	1.68	1.20	1.24	1.33
MSE	6.61	10.67	13.15	14.52	8.16	8.17	9.19
MedAE	0.23	0.38	0.94	0.50	0.25	0.32	0.39
MAPE	0.014	0.019	0.024	0.021	0.015	0.015	0.017

6. Comparison

Through the metrics, CBQR proves to be the best alternative among the six ML models, demonstrating that the trivariate copula effectively captures the dependence between variables and leverages this to predict in unsampled areas. On the other hand, SVRL, being a linear model, struggles to fit well to the nonlinear nature of the data in this case, which explains its poor performance compared to the other models.

When comparing the results obtained during the training stage with those generated in the out-of-sample estimation, significant differences are observed that highlight the behavior and generalization capacity of each model. During the training stage, the models achieve near-perfect performance, with R^2 values close to 1 for most of them (notably CBQR, DT, ET, and RF) and relatively low error metrics such as MAE and MSE. However, in the out-of-sample estimation stage, R^2 values drop significantly, especially for machine learning models like ET, DT, and SVRL, which experience a pronounced decline in predictive

capacity (R^2 ranging between 0.49 and 0.62). Conversely, CBQR and RF maintain a better relative performance, with R^2 values of 0.77 and 0.67, respectively.

Moreover, error metrics such as MAE, MSE, and MaxRE are consistently higher in the out-of-sample estimation, indicating reduced accuracy when working with data not used during training. For example, the MaxRE of CBQR increases from 0.11 in the training stage to 22.66 in the out-of-sample estimation, highlighting greater variability in extreme case predictions. This suggests potential overfitting in some models, such as DT and ET, which perform well during training but struggle to generalize effectively.

Overall, CBQR emerges as the most robust model, maintaining the best balance of R^2 (0.77) and lower error metrics in the out-of-sample estimation, followed closely by RF, which also demonstrates notable stability between both stages. These results reaffirm the suitability of CBQR as an effective alternative in scenarios with variable data, particularly when minimizing the impact of extreme values and improving predictive capacity in unsampled areas.

A comparison of the spatial distribution of the actual data and the CBQR estimate for copper recovery is shown in Figure 12.

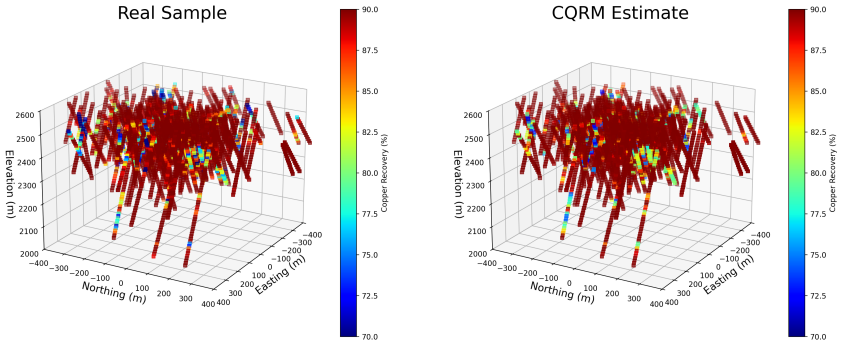

Figure 12. Real data (**left**) and CBQR for copper recovery (**right**).

7. Discussion

Conditional quantile regression, in combination with the nonparametric dependence structure of a trivariate copula, allows better adaptation to nonlinear and complex relationships between geometallurgical characteristics, thus achieving a more robust estimation in subsampling situations.

A key aspect to analyze is the use of available information and its limitation in relation to geometallurgical domains. Generally, in workflows, geometallurgical domains are first defined, and then, within these, the metallurgical response is estimated using regression methods. However, in many cases, domains lack sufficient information to build a robust predictive model. In this study, given that the relationship between CR and the predictor variables remained consistent in all geometallurgical units, the integration of complete information allowed us to improve the robustness and precision of the model.

The results obtained demonstrate that CBQR outperforms the six ML models evaluated in accuracy. The modeled dependence structure adequately captures the nonlinear relationships between the geometallurgical variables, improving the predictive capacity even in comparison with a multilayer neural network, the second most accurate model in the analysis.

Using a trivariate copula not only outperforms ML models in accuracy, but also offers a practical and flexible alternative. Unlike ML models, which require lengthy training and constant optimization of hyperparameters, copula allows direct implementation with minimal adjustments, facilitating its applicability in conditions of limited and heterogeneous data.

Although the trivariate copula model has shown acceptable accuracy in this study, its applicability may be restricted in cases where the dependence between variables is

weak. While the model effectively captures the dependency between features, it also has certain limitations:

- Dependency between predictive characteristics and variable of interest: There must be at least a weak dependence between the predictor characteristics and the target variable, and these must not be redundant, that is, they must not be correlated with each other. Otherwise, a trivariate copula does not present significant improvements over a bivariate copula [23].
- Estimation limited to locations with known data: CR can only be estimated at locations where predictive characteristics are available. For the application case, this involves deriving the estimate of CR from an imputation of data at the exploration drilling level. For the estimation of mineral resources and the economic valuation of blocks before optimization, design and mining planning, it is essential to know these predictive characteristics, in this case, the bornite and chalcocite contents at the level of blocks that represent the geometallurgical domain.

The proposed solution is to associate the error of the CR estimate with the error of the geostatistical estimate of the predictor variables. Both bornite and chalcocite are additive mineralogical variables that can be geostatistically modeled and estimated using a linear, unbiased and minimum variance method, widely validated in the mining industry as is the case of Ordinary Kriging.

To do this, a variogram model and an independent estimation plan are prepared for each of these variables. Subsequently, with the results of both estimates and the trivariate copula model previously calibrated at the sample level, the median quantile regression provides the estimate of the expected metallurgical recovery. Figure 13 compares the 3D maps of the spatial distribution of bornite and chalcocite estimated by Ordinary Kriging and the CR estimation by CBQR conditioned on the previously estimated variables.

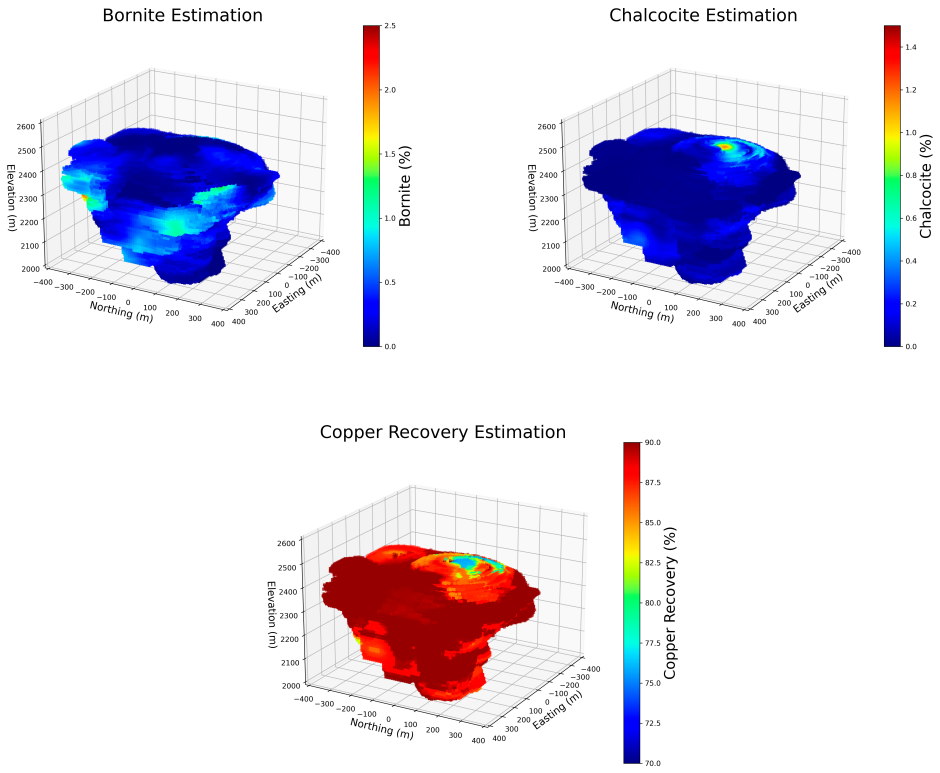

Figure 13. Three-dimensional maps of the spatial distribution of bornite and chalcocite (**top**) estimated using Ordinary Kriging and the CR estimation using CBQR (**bottom**) conditioned on the previously estimated variables.

8. Conclusions

The combination of the multivariate copula model with conditional quantile regression proved to be an effective methodological approach for modeling complex multivariable dependencies in scenarios characterized by asymmetric distributions, nonlinear relationships, and significantly undersampled data. The application case results show that this approach outperforms the six supervised learning models evaluated under the same input conditions. It stands out with a coefficient of determination of $R^2 = 0.77$, surpassing models such as SVRE and MLP ($R^2 = 0.71$) or DT ($R^2 = 0.62$). Additionally, it achieves the lowest values across metrics, including MAE = 1.11, MaxRE = 22.66, MSE = 6.61, MedAE = 0.23, and MAPE = 0.014. These results establish the CBQR method as a robust and efficient alternative compared to the supervised methods analyzed.

The model's ability to capture complex relationships and estimate specific quantiles of metallurgical recovery underscores its utility in the field of geometallurgy, particularly in applications where data are scarce and costly to obtain. Moreover, the model demonstrates adaptability by integrating with geostatistical methods, such as Kriging, to extend its applicability to unsampled areas while inheriting errors from primary variables. Despite certain limitations, such as the requirement for a minimum level of dependency between predictor variables and the target variable, this approach offers an innovative and generalizable solution for addressing challenges in mining and other disciplines. Its capability to model complex data positions it as a significant contribution to advancing data analysis and modeling in challenging contexts.

Finally, it is worth mentioning that this research has awakened the interest of several companies, e.g., the steel company Sidenor, as it is continuously implementing skilling and reskilling programs for the integration of these and other data-driven decision-making tools.

Author Contributions: Conceptualization, H.H., M.A.D.-V., and A.G.; investigation, H.H., E.A., and A.G.; methodology, M.A.D.-V. and E.A.; software, M.A.D.-V.; validation, H.H. and E.A.; formal analysis, M.A.D.-V.; writing—original draft preparation, H.H. and M.A.D.-V.; writing—review and editing, E.A. and A.G.; supervision, E.A.; funding acquisition, A.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by project SUSTASKILLS: Development of a roadmap for the implementation of skills related to industrial symbiosis and energy efficiency to achieve a sustainable process industry (Grant Agreement No PUE_2023_1_0006). The sole responsibility for the issues treated in the present paper lies with the authors; the Commission is not responsible for any use that may be made of the information contained therein.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CR	Copper Recovery
CBQR	Copula-Based Conditional Quantile Regression
DT	Decision Tree
ET	Extra Trees
SVRL	Support Vector Regression (linear)
SVRE	Support Vector Regression (with epsilon)
MLP	Multi-Layer Perceptron
RF	Random Forest

R^2	Determination coefficient
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MedAE	Median Absolute Error
MaxRE	Maximum Relative Error
ML	Machine Learning
MSE	Mean Squared Error
PDF	Probability Density Function

References

1. Afum, B.; Ben-Awuah, E. A Review of Models and Algorithms for Surface-Underground Mining Options and Transitions Optimization: Some Lessons Learnt and the Way Forward. *Mining* **2021**, *1*, 112–134. [CrossRef]
2. Jara, R.; Couble, A.; Emery, X.; Magri, E.; Ortiz, J. Block size selection and its impact on open-pit design and mine planning. *J. S. Afr. Inst. Min. Metall.* **2006**, *106*, 205–211.
3. Letelier, O.; Espinoza, D.; Goycoolea, M.; Moreno, E.; Muñoz, G. Production scheduling for strategic open pit mine planning: A mixed-integer programming approach. *Oper. Res.* **2020**, *68*, 1425–1444. [CrossRef]
4. Mai, N.; Topal, E.; Erten, O. A new open-pit mine planning optimization method using block aggregation and integer programming. *J. S. Afr. Inst. Min. Metall.* **2018**, *118*, 705–714. [CrossRef]
5. Pell, R.; Tijsseling, L.; Palmer, L.; Glass, H.; Yan, X.; Wall, F.; Zeng, X.; Li, J. Environmental optimisation of mine scheduling through life cycle assessment integration. *Resour. Conserv. Recycl.* **2019**, *142*, 267–276. [CrossRef]
6. Morales, N.; Seguel, S.; Cáceres, A.; Jélvez, E.; Alarcón, M. Incorporation of geometallurgical attributes and geological uncertainty into long-term open-pit mine planning. *Minerals* **2019**, *9*, 108. [CrossRef]
7. Becerra, M.; Jerez, A.; Garcés, H.; Demarco, R. Copper price: A brief analysis of China's impact over its short-term forecasting. *Resour. Policy* **2022**, *75*, 102449. [CrossRef]
8. Yushan, P.; Ni, M.; Wang, X. Identifying price bubbles in copper market: Evidence from a GSADF test approach. *PLoS ONE* **2023**, *18*, e0290983. [CrossRef]
9. Alberdi, E.; Hernández, H.; Goti, A. Development of methods based on neural networks in the estimation of mineral resources. *Dyna* **2024**, *99*, 303–310. [CrossRef]
10. Chilès, J.; Desassis, N. Fifty years of kriging. In *Handbook of Mathematical Geosciences: Fifty Years of IAMG*; Springer: Cham, Switzerland, 2018; pp. 589–612. [CrossRef]
11. Remy, N.; Boucher, A.; Wu, J. *Applied Geostatistics with SGeMS*; Cambridge University Press: Cambridge, UK, 2009. [CrossRef]
12. Henckens, M.; van Ierland, E.; Driessen, P.; Worrell, E. Mineral resources: Geological scarcity, market price trends, and future generations. *Resour. Policy* **2016**, *49*, 102–111. [CrossRef]
13. Dominy, S.; O'Connor, L.; Parbhakar-Fox, A.; Glass, H.; Purevgerel, S. Geometallurgy—A route to more resilient mine operations. *Minerals* **2018**, *8*, 560. [CrossRef]
14. Mu, Y.; Salas, J. Data-Driven Synthesis of a Geometallurgical Model for a Copper Deposit. *Processes* **2023**, *11*, 1775. [CrossRef]
15. van den Boogaart, K.G.; Tolosana-Delgado, R. Predictive geometallurgy: An interdisciplinary key challenge for mathematical geosciences. In *Handbook of Mathematical Geosciences: Fifty Years of IAMG*; Springer: Cham, Switzerland, 2018; pp. 673–686. [CrossRef]
16. Abildin, Y.; Madani, N.; Topal, E. A hybrid approach for joint simulation of geometallurgical variables with inequality constraint. *Minerals* **2019**, *9*, 24. [CrossRef]
17. Boisvert, J.; Rossi, M.; Ehrig, K.; Deutsch, C. Geometallurgical Modeling at Olympic Dam Mine, South Australia. *Math. Geosci.* **2023**, *45*, 901–925. [CrossRef]
18. Both, C.; Dimitrakopoulos, R. Geometallurgical prediction models of processing plant indicators for stochastic mine production scheduling. *IFAC-PapersOnLine* **2022**, *55*, 1267. [CrossRef]
19. Deutsch, J.; Palmer, K.; Deutsch, C.; Szymanski, J.; Etsell, T. Spatial Modeling of Geometallurgical Properties: Techniques and a Case Study. *Nat. Resour. Res.* **2016**, *25*, 161–181. [CrossRef]
20. Deutsch, J.; Szymanski, J.; Etsell, T. Metallurgical variable re-expression for geostatistics. In *Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the Environment: Challenges, Processes and Strategies*; Springer International Publishing: Berlin, Germany, 2016; pp. 83–88. [CrossRef]
21. Garrido, M.; Ortiz, J.; Villaseca, F.; Kracht, W.; Townley, B.; Miranda, R. Change of support using non-additive variables with Gibbs Sampler: Application to metallurgical recovery of sulphide ores. *Comput. Geosci.* **2019**, *122*, 68–76. [CrossRef]
22. Garrido, M.; Sepúlveda, E.; Ortiz, J.; Townley, B. Simulation of Synthetic Exploration and Geometallurgical Database of Porphyry Copper Deposits for Educational Purposes. *Nat. Resour. Res.* **2020**, *29*, 3527–3545. [CrossRef]

23. Hernández, H.; Díaz-Viera, M.; Alberdi, E.; Oyarbide-Zubillaga, A.; Goti, A. Metallurgical Copper Recovery Prediction Using Conditional Quantile Regression Based on a Copula Model. *Minerals* **2024**, *14*, 691. [CrossRef]
24. Hosseini, S.; Asghari, O. Simulation of geometallurgical variables through stepwise conditional transformation in Sungun copper deposit, Iran. *Arab. J. Geosci.* **2015**, *8*, 3821–3831. [CrossRef]
25. Hunt, J.; Berry, R. Economic geology models 3. Geological contributions to geometallurgy: A review. *Geosci. Can.* **2017**, *44*, 103–118. [CrossRef]
26. Van Tonder, E.; Deglon, D.; Napier-Munn, T. The effect of ore blends on the mineral processing of platinum ores. *Miner. Eng.* **2010**, *23*, 621–626. [CrossRef]
27. Bhuiyan, M.; Esmaieli, K.; Ordóñez-Calderón, J. Application of data analytics techniques to establish geometallurgical relationships to bond work index at the Paracutu Mine, Minas Gerais, Brazil. *Minerals* **2019**, *9*, 302. [CrossRef]
28. Dachri, K.; Bouabidi, M.; Naji, K.; Nouar, K.; Benzakour, I.; Oummouch, A.; Hibti, M.; El Amari, K. Predictive insights for copper recovery: A synergistic approach integrating variability data and machine learning in the geometallurgical study of the Tizert deposit, Morocco. *J. Afr. Earth Sci.* **2024**, *212*, 105208. [CrossRef]
29. Fouedjio, F.; Hill, E.; Laukamp, C. Geostatistical clustering as an aid for ore body domaining: Case study at the Rocklea Dome channel iron ore deposit, Western Australia. *Appl. Earth Sci. Trans. Inst. Min. Metall.* **2018**, *127*, 15–29. [CrossRef]
30. Tiu, G.; Ghorbani, Y.; Jansson, N.; Wanhainen, C.; Bolin, N. Quantifying the variability of a complex ore using geometallurgical domains. *Miner. Eng.* **2023**, *203*, 108323. [CrossRef]
31. Bruckard, W.; Sparrow, G.; Woodcock, J. A review of the effects of the grinding environment on the flotation of copper sulphides. *Int. J. Miner. Process.* **2011**, *100*, 1–13. [CrossRef]
32. Lee, K.; Archibald, D.; McLean, J.; Reuter, M. Flotation of mixed copper oxide and sulphide minerals with xanthate and hydroxamate collectors. *Miner. Eng.* **2009**, *22*, 395–401. [CrossRef]
33. Hoffimann, J.; Augusto, J.; Resende, L.; Mathias, M.; Mazzinghy, D.; Bianchetti, M.; Mendes, M.; Souza, T.; Andrade, V.; Domingues, T.; et al. Modeling Geospatial Uncertainty of Geometallurgical Variables with Bayesian Models and Hilbert–Kriging. *Math. Geosci.* **2022**, *54*, 1227–1253. [CrossRef]
34. Carrasco, P.; Chilès, J.; Séguret, S.A. Additivity, Metallurgical Recovery, and Grade. In Proceedings of the 8th International Geostatistics Congress, Santiago, Chile, 1–5 December 2008; p. 1188.
35. Lishchuk, V.; Koch, P.; Ghorbani, Y.; Butcher, A. Towards integrated geometallurgical approach: Critical review of current practices and future trends. *Miner. Eng.* **2020**, *145*, 106072. [CrossRef]
36. McCoy, J.; Auret, L. Machine learning applications in minerals processing: A review. *Miner. Eng.* **2019**, *132*, 95–109. [CrossRef]
37. Sepulveda, E.; Dowd, P.; Xu, C.; Addo, E. Multivariate Modelling of Geometallurgical Variables by Projection Pursuit. *Math. Geosci.* **2017**, *49*, 121–143. [CrossRef]
38. Ortiz, J.; Kracht, W.; Townley, B.; Lois, P.; Cardenas, E.; Miranda, R.; Alvarez, M. Workflows in geometallurgical prediction: Challenges and outlook. In Proceedings of the IAMG 2015–17th Annual Conference of the International Association for Mathematical Geosciences, Freiberg, Germany, 5–13 September 2015; pp. 228–234.
39. Rajabinasab, B.; Asghari, O. Geometallurgical Domaining by Cluster Analysis: Iron Ore Deposit Case Study. *Nat. Resour. Res.* **2019**, *28*, 665–684. [CrossRef]
40. Sepúlveda, E.; Dowd, P.; Xu, C. Fuzzy Clustering with Spatial Correction and Its Application to Geometallurgical Domaining. *Math. Geosci.* **2018**, *50*, 895–928. [CrossRef]
41. Madenova, Y.; Madani, N. Application of Gaussian Mixture Model and Geostatistical Co-simulation for Resource Modeling of Geometallurgical Variables. *Nat. Resour. Res.* **2021**, *30*, 1199–1228. [CrossRef]
42. Both, C.; Dimitrakopoulos, R. Applied machine learning for geometallurgical throughput prediction—A case study using production data at the tropicana gold mining complex. *Minerals* **2021**, *11*, 1257. [CrossRef]
43. Durante, F.; Puccetti, G.; Scherer, M.; Vanduffel, S. My introduction to copulas. *Depend. Model.* **2017**, *5*, 88–98. [CrossRef]
44. Genest, C.; Rivest, L. Statistical inference procedures for bivariate Archimedean copulas. *J. Am. Stat. Assoc.* **1993**, *88*, 1034–1043. [CrossRef]
45. Joe, H. *Dependence Modeling with Copulas*; CRC Press: Vancouver, BC, Canada, 2014. [CrossRef]
46. Trivedi, P.; Zimmer, D. Copula modeling: An introduction for practitioners. *Found. Trends Econom.* **2005**, *1*, 1–111. [CrossRef]
47. Frenzel, M.; Baumgartner, R.; Tolosana-Delgado, R.; Gutzmer, J. Geometallurgy: Present and Future. *Elements* **2023**, *19*, 345–351. [CrossRef]
48. Suazo, C.; Kracht, W.; Alruiz, O. Geometallurgical modelling of the Collahuasi flotation circuit. *Miner. Eng.* **2010**, *23*, 137–142. [CrossRef]
49. Adugna, T.D.; Ramu, A.; Haldorai, A. A Review of Pattern Recognition and Machine Learning. *J. Mach. Comput.* **2024**, *4*, 210–220. [CrossRef]
50. Heaton, J.; Goodfellow, I.; Bengio, Y.; Courville, A. Deep learning. *Genet. Program. Evolvable Mach.* **2018**, *19*, 305–307. [CrossRef]
51. Lai, C.; Balakrishna, N. *Continuous Bivariate Distributions*; Springer: New York, NY, USA, 2009. [CrossRef]

52. Jaworski, P.; Fabrizio, D.; Wolfgang, R.; HärdleTomasz, K. Copula Theory and Its Applications. In *Copula Theory and Its Applications*; Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., Zeger, S., Eds.; Springer: Berlin, Germany, 2010; p. 198. [CrossRef]
53. Nelsen, R. Properties and applications of copulas: A brief survey. In Proceedings of the Proceedings of the first Brazilian conference on statistical modeling in insurance and finance, Sao Paulo, Brazil, 1–6 September 2003; pp. 10–28.
54. Sklar, M. Fonctions de répartition à n dimensions et leurs marges. *Ann. De L'ISUP* **1959**, *8*, 229–231.
55. Nelsen, R. *An Introduction to Copulas*, 2nd ed.; Springer: New York, NY, USA, 2006. [CrossRef]
56. Baudin, M.; Dutfoy, A.; Iooss, B.; Popelin, A. OpenTURNS: An Industrial Software for Uncertainty Quantification in Simulation. In *Handbook of Uncertainty Quantification*; Ghanem, R., Higdon, D., Owhadi, H., Eds.; Springer: Cham, Switzerland, 2015; pp. 1–38. [CrossRef]
57. Song, Y.Y.; Lu, Y. Decision tree methods: Applications for classification and prediction. *Shanghai Arch. Psychiatry* **2015**, *27*, 130. [CrossRef] [PubMed]
58. Zhang, F.; O'Donnell, L.J. Chapter 7—Support vector regression. In *Machine Learning*; Mechelli, A., Vieira, S., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 123–140. [CrossRef]
59. Du, K.L.; Leung, C.S.; Mow, W.H.; Swamy, M.N.S. Perceptron: Learning, Generalization, Model Selection, Fault Tolerance, and Role in the Deep Learning Era. *Mathematics* **2022**, *10*, 4730. [CrossRef]
60. Schonlau, M.; Zou, R.Y. The random forest algorithm for statistical learning. *Stata J.* **2020**, *20*, 3–29. [CrossRef]
61. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
62. Käyhkö, T.; Sinche-Gonzalez, M.; Khizanishvili, S.; Liipo, J. Validation of predictive flotation models in blended ores for concentrator process design. *Miner. Eng.* **2022**, *185*, 107685. [CrossRef]