Math Geosci (2022) 54:1227–1253
https://doi.org/10.1007/s11004-022-10013-1

Modeling Geospatial Uncertainty of Geometallurgical Variables with Bayesian Models and Hilbert–Kriging

Júlio Hoffimann[1] · **José Augusto**[1] · **Lucas Resende**[1] · **Marlon Mathias**[1] · **Douglas Mazzinghy**[2] · **Matheus Bianchetti**[3] · **Mônica Mendes**[3] · **Thiago Souza**[3] · **Vitor Andrade**[3] · **Tarcísio Domingues**[3] · **Wesley Silva**[3] · **Ruberlan Silva**[3] · **Danielly Couto**[3] · **Elisabeth Fonseca**[3] · **Keila Gonçalves**[3]

Received: 8 March 2022 / Accepted: 4 July 2022 / Published online: 12 August 2022

Abstract In mine planning, geospatial estimates of variables such as comminution indexes and metallurgical recovery are extremely important to locate blocks for which the energy consumption at the plant is minimized and for which the recovery of minerals is maximized. Unlike ore grades, these variables cannot be modeled with traditional geostatistical methods, which rely on the availability of a large number of samples for variogram estimation and on the additivity of variables for change of support, among other issues. Past attempts to build geospatial models of geometallurgical variables have failed to address some of these issues, and most importantly, did not consider adequate mathematical models for uncertainty quantification. In this work, we propose a new methodology that combines Bayesian predictive models with Kriging in Hilbert spaces to quantify the geospatial uncertainty of such variables in realistic industrial settings. The results we obtained with data from a real deposit indicate that the proposed approach may become an interesting alternative to geostatistical simulation.

Keywords Bayesian modeling · Kriging · Hilbert spaces · Drop weight test · Bond work index · Metallurgical recovery · Geostatistics · Geometallurgy

✉ Júlio Hoffimann
julio.hoffimann@impa.br

1 Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil

2 Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

3 Vale S.A., Rio de Janeiro, Brazil

1 Introduction

In the mining industry, the exploitation of mineral resources is planned on the basis of geospatial models of ore grades, and more recently, geometallurgical variables such as comminution indexes [e.g., drop weight test (DWT), Bond work index (BWI)] and metallurgical recovery. Laboratory experiments to estimate geometallurgical variables are time-consuming compared to the chemical analysis that is performed meter by meter along drill holes, and consequently, only a few (< 100) core samples of greater support (e.g., 10 meters) are subject to comminution and flotation tests. Moreover, the distributions of these variables cannot be Gaussian (e.g., bounded support). In particular, it is well known that comminution indexes and metallurgical recovery are not additive—the component-weighted average of two sample values is not a good estimator of the corresponding value in the blend (Tavares and Kallemback 2013; Carrasco et al. 2008)—and that this property hinders the use of linear geostatistical models. Ignoring these issues can lead to bias in geospatial estimates and suboptimal mine planning (Campos et al. 2021).

Although geostatistical methods such as Kriging and Gaussian simulation have been extensively proposed for ore grades (Journel 2003), these same methods cannot be directly applied to geometallurgical variables for various reasons, which include a reduced number of samples for variogram estimation, non-additive variables, and non-Gaussian distributions, among others. Past attempts to build geospatial models of geometallurgical variables did not address these issues or considered modeling assumptions that are not valid in industrial settings.

A common strategy to circumvent the modeling challenges in geometallurgy consists of (1) applying nonlinear transformations to the data, (2) modeling the data as if they were Gaussian, and (3) undoing the transformations (Boisvert et al. 2013; Deutsch et al. 2015). In spite of its appeal, this strategy cannot handle physical constraints that are crucial in downstream applications (e.g., material balance, positivity of hardness), nor can it quantify uncertainty adequately. Consequently, models created with this strategy usually require detailed human intervention and ad hoc modifications to work in practice. They are not robust to small variations in the data, cannot be applied online, and cannot be easily transferred across different types of deposits.

For specific deposits, it is sometimes possible to exploit relationships between a set of auxiliary variables to design geospatial models of primary geometallurgical variables. As an example, consider the metallurgical recovery of copper, defined as the mass ratio of copper in the concentrate by copper in the feed in a flotation process. If the numerator and denominator are linearly related, then it is possible to use co-Kriging to estimate both (additive) variables simultaneously (assuming enough samples are available for variogram estimation), and consequently estimate the ratio at each mining block (Adeli et al. 2021). Although clever, this model is not general enough for wider use by the industry, which is usually concerned with nonlinear recoveries from locked cycle tests.

Finally, and most importantly, the misspecification of Gaussian distributions for geometallurgical (or transformed) variables may lead to poor uncertainty estimates. Together with the reduced number of samples that is common in geometallurgical modeling, this is extremely undesirable.

In this work, we propose a new methodology that combines Bayesian predictive models (Gelman 2014) with Kriging in Hilbert spaces (Hilbert–Kriging) (Menafoglio et al. 2013) to quantify the geospatial uncertainty of geometallurgical variables. Bayesian models allow for the incorporation of domain expertise in the form of prior distributions and physical relations, which counterweights the reduced number of samples. Hilbert–Kriging enables the interpolation of non-Gaussian distributions from drill holes to mining blocks. As a result, the proposed methodology provides quick ($\approx$ 6 min) probabilistic estimates of non-Gaussian variables across space without compute-intensive geostatistical simulation.

The paper is organized as follows. In Sect. 2, we define Bayesian models for different geometallurgical variables and explain how the output of these models is used in Hilbert–Kriging. All the steps of the proposed methodology are illustrated using data from a real copper deposit. In Sect. 3, we discuss the results and possible technical challenges associated with the application of the methodology in practice. In Sect. 4, we conclude the work and point to future research directions.

2 Methodology

2.1 Overview

The flowchart in Fig. 1 illustrates the steps of the proposed methodology, which starts with three tables: `comminution`, `flotation`, and `drillholes`. The `comminution` and `flotation` tables contain data for a reduced number of rock samples—cylinders—that are subject to comminution and flotation tests. The content of these tables will be described in more detail in the following sections. The `drillholes` table is the standard table in the mining industry that contains chemical analysis data for all rock samples along the drill holes with their respective X, Y, and Z coordinates.

In the first step, the `comminution` and `flotation` tables are used to infer the posterior distribution of geometallurgical variables according to a set of Bayesian models. These models take chemical (or mineralogy) data as input and output full probability distributions for the drop weight test (DWT), Bond work index (BWI), and metallurgical recovery of copper in a locked cycle test (LCT). By assuming that the samples in the `drillholes` table are composited to a length that is similar to the length of the samples in the other two tables, the Bayesian models can be directly applied to the rock samples in the drill holes. The result of this second step consists of a new `predictions` table that contains probabilistic predictions for all rock samples along the drill holes. In the final step, the probability distributions of the DWT, BWI, and LCT are interpolated for all blocks with Hilbert–Kriging, which accounts for the change of support. The final result is a block model where each block has a full probability distribution for each geometallurgical variable. No geoestatistical simulation is involved.

In Fig. 2, we illustrate the geospatial configuration of the samples from the real copper deposit investigated in this work. The nature and number of `comminution` and `flotation` samples invalidate the use of traditional geostatistical methods.

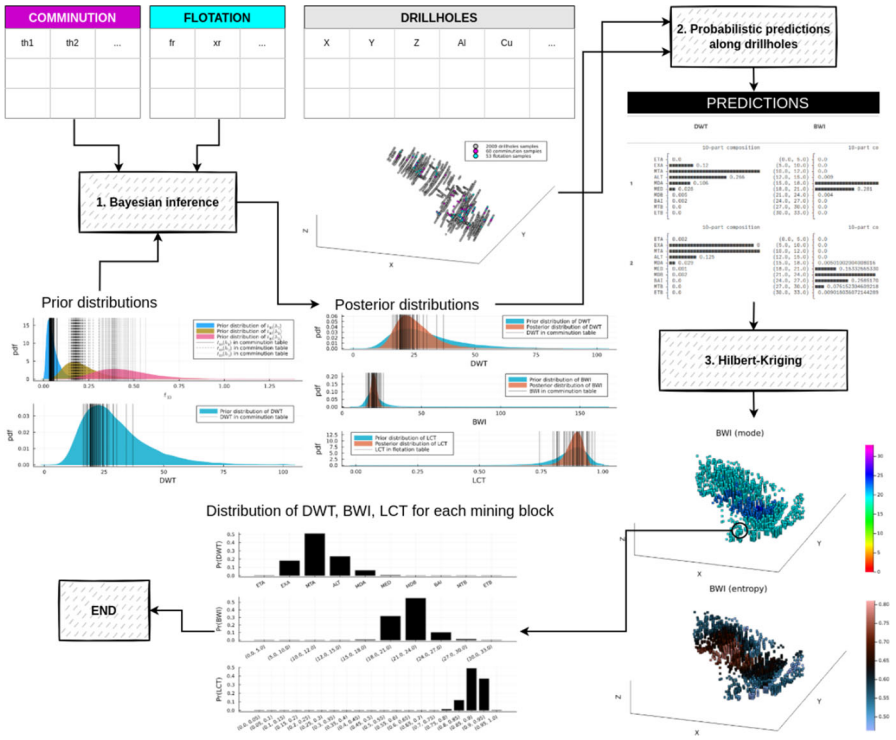


Fig. 1 Flowchart of the proposed methodology

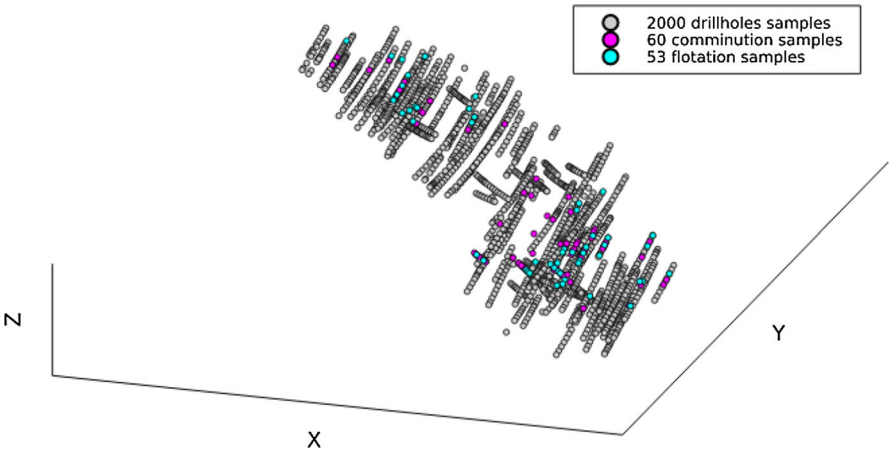


Fig. 2 Geospatial configuration of `comminution`, `flotation`, and `drillholes` samples in a real copper deposit

2.2 Bayesian Models

Hereafter, we assume that the reader is familiar with Bayesian statistics. The excellent introductory book by Davidson-Pilon (2015) is freely available online for those who are not familiar with basic concepts such as Bayes' rule and prior distributions.

The following sections describe three specific Bayesian models for DWT, BWI, and LCT, but other Bayesian models could have been adopted instead. Each model takes a vector $\boldsymbol{x} \in \mathbb{R}^p$ of explanatory variables as input and outputs probabilistic predictions of the response geometallurgical variable. In this work, the explanatory variables are chemical compositions, which are also available in the `drillholes` table. In order to facilitate the specification of prior distributions and to reduce the effects of abundant elements, the explanatory variables are transformed with a centered log-ratio (CLR) (Aitchison 2003) followed by a normal quantile (Barnett and Deutsch 2012) (a.k.a. normal score) transform.

2.2.1 Drop Weight Test

In this comminution test, a standard weight is dropped from three heights h_1, h_2, h_3 on n rock samples, and particle size distributions are recorded after each impact $\phi_{h_j}^{(i)},\ i = 1, 2, \ldots, n,\ j = 1, 2, 3$. For each height h_j, there is a corresponding potential energy E_{h_j}, which is considered an input to the test. The value $t_{10}^{(i)}(h_j) = \phi_{h_j}^{(i)}(l/10)$ is defined as the fraction of particles with size smaller or equal to $l/10$, where l is the original size of each and every sample (i.e., $l^{(1)} = l^{(2)} = \cdots = l^{(n)} = l$).

The standardized formula

$$t_{10}(h) = A\left(1 - e^{-bE_h}\right) \tag{1}$$

relates the potential energy to the fraction of fine material (Napier-Munn et al. 1999). The parameters $A > 0$ and $b > 0$ are fitted for each sample with least squares, considering the three energy levels. Let $J_i(A, b) = \sum_{j=1}^{3}\left(t_{10}^{(i)}(h_j) - A\left(1 - e^{-bE_{h_j}}\right)\right)^2$ be the sum of squares for the i-th sample. It can be written in matrix form as

$$J_i(A, b) = \left\| \boldsymbol{t}^{(i)} - A\left(\mathbf{1} - e^{-b\boldsymbol{E}}\right)\right\|_2^2, \tag{2}$$

with $\boldsymbol{E} = (E_{h_1}, E_{h_2}, E_{h_3})$ and $\boldsymbol{t}^{(i)} = (t_{10}^{(i)}(h_1), t_{10}^{(i)}(h_2), t_{10}^{(i)}(h_3))$. The energy levels, which are also standardized in the industry, are shared across all samples. The optimal solution is obtained via minimization,

$$A^{(i)}, b^{(i)} = \underset{A,b>0}{\arg\min}\, J_i(A, b). \tag{3}$$

Finally, the drop weight test index is defined as $\mathrm{DWT}^{(i)} = A^{(i)} \times b^{(i)}$. It summarizes the fragility of the sample in the following sense: the higher the index, the more fragile the sample.

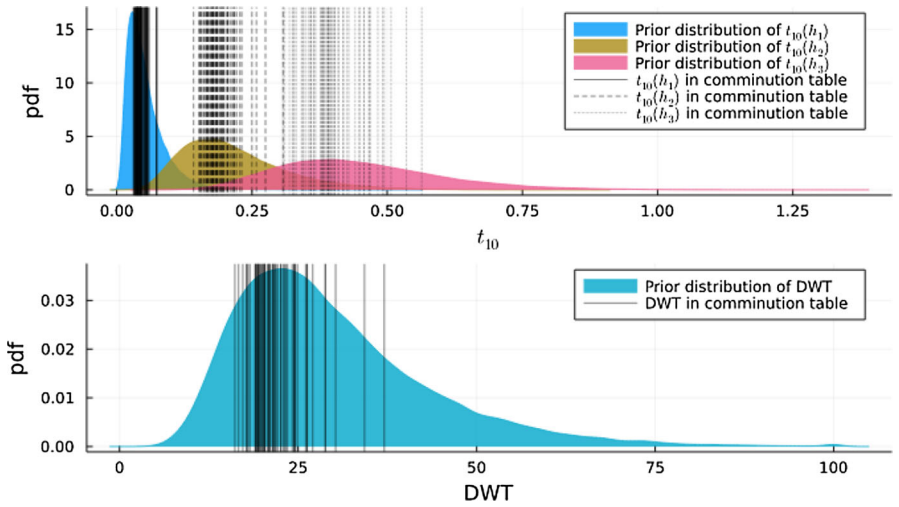


Fig. 3 Prior distribution of $t_{10}(h_1)$, $t_{10}(h_2)$, $t_{10}(h_3)$, and consequently of DWT. Data from the `comminution` table illustrated with vertical lines

In Appendix A, we demonstrate that the DWT index cannot be a linear function of any given set of explanatory variables. In particular, it cannot be modeled with linear geostatistical models. Motivated by this fact, and aware of the variables involved in the computation of the index, we propose the following Bayesian hierarchical model,

$$t_{10}(h_j) = t_{10}(h_{j-1}) + \Delta(h_j) \tag{4}$$

$$\Delta(h_j) \sim \text{LogitNormal}(\mu_j,\ \sigma_j^2) \tag{5}$$

$$\mu_j = \alpha_j + \langle \boldsymbol{\beta}_j, \boldsymbol{x} \rangle \tag{6}$$

$$\alpha_j \sim \text{Normal}(\delta_j,\ \sigma_j^2) \tag{7}$$

$$\boldsymbol{\beta}_j \sim \text{Normal}(\mathbf{0},\ \tau_j^2 \boldsymbol{I}). \tag{8}$$

The model defines random variables for the fractions of fine material in terms of increments $\Delta(h_j) = t_{10}(h_j) - t_{10}(h_{j-1})$. These increments are random variables themselves with a LogitNormal(μ_j, σ_j^2) distribution. The mean parameter $\mu_j = \alpha_j + \langle \boldsymbol{\beta}_j, \boldsymbol{x} \rangle$ of this distribution is assumed to be an affine combination of explanatory variables with normally distributed coefficients $\boldsymbol{\beta}_j$ and intercept α_j. By fixing the hyperparameters of the model $\delta_j, \sigma_j, \tau_j$, we can sample the prior distribution of fractions, and consequently, the prior distribution of DWT as illustrated in Fig. 3.

The hyperparameters δ_j locate the centers of the three density plots. They are initialized with empirical averages from the `comminution` table,

$$\delta_1 = \frac{1}{n}\sum_{i=1}^{n} t_{10}^{(i)}(h_1) \tag{9}$$

$$\delta_2 = \frac{1}{n}\sum_{i=1}^{n} t_{10}^{(i)}(h_2) - t_{10}^{(i)}(h_1) \tag{10}$$

$$\delta_3 = \frac{1}{n}\sum_{i=1}^{n} t_{10}^{(i)}(h_3) - t_{10}^{(i)}(h_2) \tag{11}$$

The hyperparameters σ_j and τ_j specify the prior uncertainty around these centers. They are initialized with fixed large values to encompass all possible physical values. All hyperparameters can be adjusted by the domain expert if he/she believes that the default prior distribution for the deposit is different.

2.2.2 Bond Work Index

The Bond work index (BWI) in kWh/t is an index that quantifies the resistance of a sample to grinding according to a test carried out in a laboratory mill. The laboratory mill has an internal diameter of $305mm$ and length of $305mm$. The mill has a smooth lining withrounded corners and no lifters. The ball media charge has approximately $20.1kg$ distributed in different sizes starting with $38mm$ (Bond 1961). Ultimately, the BWI is estimated using the formula

$$\text{BWI} = \frac{49.0}{A^{0.23}M^{0.82}\left(\frac{10.0}{\sqrt{P80}} - \frac{10.0}{\sqrt{F80}}\right)}, \tag{12}$$

where $F80$ and $P80$ are the apertures in μm for which 80% of the material in the feed and in the product passes through; M is the moability index in g/rev; and A is the size of the screen used in μm. Equation 12 was derived by Bond decades ago. Even though its exponents were calibrated around that time with a specific data set, the formula is still widely used in the industry for a variety of mineral deposits, possibly adopting correction factors for specific cases (Rowland 1975). In this work, we adopt the formula as is, but emphasize the need for a full data-driven approach without preset constants.

To improve the convergence of Bayesian inference in Sect. 2.3 and to facilitate the specification of prior distributions, we normalize the variables as follows. Let $m_F = \frac{1}{n}\sum_{i=1}^{n} F80^{(i)}$ be the empirical average of $F80$ in the `comminution` table, and let $G = P80/F80$ be the ratio of apertures. For a fixed screen size A, we propose the following Bayesian hierarchical model,

$$\frac{F80}{m_F} \sim \text{Normal}(\mu_F, \sigma_F^2) \tag{13}$$

$$G \sim \text{LogitNormal}(\mu_G, \sigma_G^2) \tag{14}$$

$$\frac{P80}{m_F} = G\frac{F80}{m_F} \tag{15}$$

$$M \sim \text{Normal}(\mu_M, \sigma_M^2) \tag{16}$$
$$\mu_F = \alpha_F + \langle \boldsymbol{\beta}_F, \boldsymbol{x} \rangle \tag{17}$$
$$\mu_G = \alpha_G + \langle \boldsymbol{\beta}_G, \boldsymbol{x} \rangle \tag{18}$$
$$\mu_M = \alpha_M + \langle \boldsymbol{\beta}_M, \boldsymbol{x} \rangle \tag{19}$$
$$\alpha_F \sim \text{Normal}(1, \sigma_F^2) \tag{20}$$
$$\alpha_G \sim \text{Normal}(m_G, \sigma_G^2) \tag{21}$$
$$\alpha_M \sim \text{Normal}(m_M, \sigma_M^2) \tag{22}$$
$$\boldsymbol{\beta}_F \sim \text{Normal}(\mathbf{0}, \tau_F^2 \boldsymbol{I}) \tag{23}$$
$$\boldsymbol{\beta}_G \sim \text{Normal}(\mathbf{0}, \tau_G^2 \boldsymbol{I}) \tag{24}$$
$$\boldsymbol{\beta}_M \sim \text{Normal}(\mathbf{0}, \tau_M^2 \boldsymbol{I}). \tag{25}$$

Due to the normalization by m_F, the distribution of α_F is centered at 1. The hyperparameters m_G and m_M locate the center of the other two density plots in Fig. 4. They are initialized with empirical averages from the `comminution` table

$$m_G = \frac{1}{n} \sum_{i=1}^{n} \frac{P80^{(i)}}{F80^{(i)}} \tag{26}$$
$$m_M = \frac{1}{n} \sum_{i=1}^{n} M^{(i)}. \tag{27}$$

Similar to the DWT model, the hyperparameters σ_F , σ_G, σ_M, and τ_F , τ_G, τ_M specify the prior uncertainty around these centers. They are initialized with fixed large values. By fixing the hyperparameters of the model, we can sample the prior distribution of $F80$, G, $P80$, M, and consequently of BWI, using Eq. 12.

2.2.3 Locked Cycle Test

A locked cycle test (LCT) is a low-cost test to estimate the metallurgical recovery of a given ore in an industrial-scale flotation circuit (Agar 2000). The LCT carried out in this project considered an arrangement of cells that are known as "rougher" and "cleaner" to separate the mass of copper sulfides from the gangue minerals that are present in the rock sample, see Fig. 5.

In a first stage, a sample of mass m is fed into the rougher cell with a known chemical composition $\boldsymbol{x} = (x^{Cu}, x^{Au}, \ldots, x^{Fe})$. A fraction f_r of this mass is recovered in the rougher concentrate with an enriched grade of copper $x_r^{Cu} \geq x^{Cu}$. In a second stage, the mass in the rougher concentrate m_r is fed into the cleaner cell (or a sequence of such cells), and a new fraction f_c is recovered with an even richer grade $x_c^{Cu} \geq x_r^{Cu}$. The metallurgical recovery of copper at the end of the second stage is defined as the mass ratio of copper in the cleaner concentrate by copper in the feed

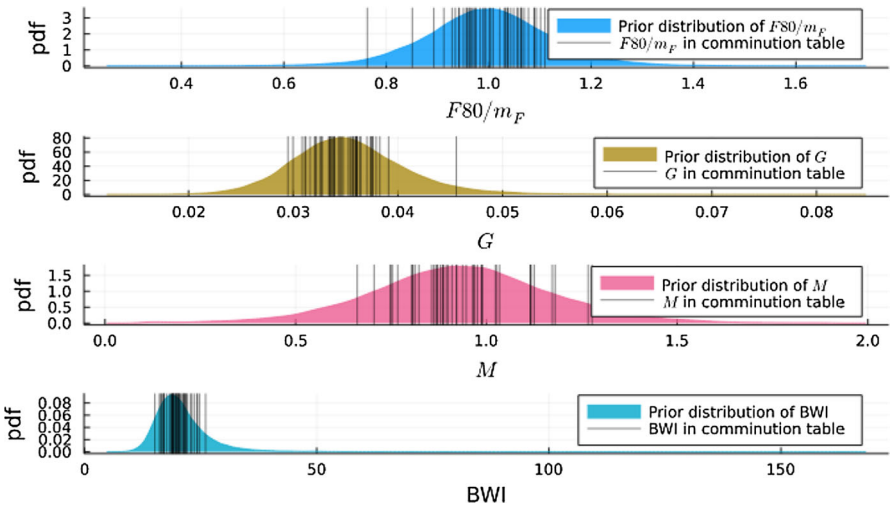


Fig. 4 Prior distribution of $F80$, G, M, and consequently of BWI. Data from the `comminution` table illustrated with vertical lines

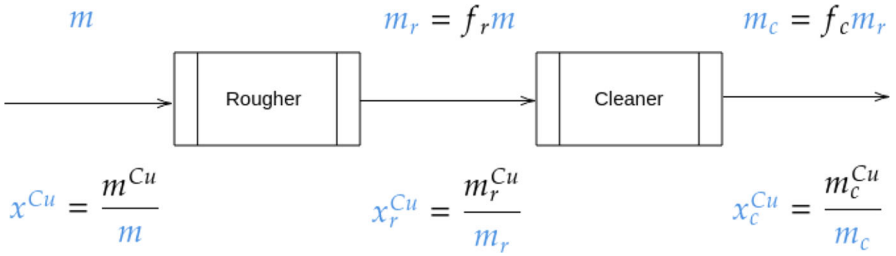


Fig. 5 Schematic illustration of "rougher-cleaner" cells inside a locked cycle test. Data from `flotation` table shown in blue

$$R_{rc}^{Cu} = \frac{m_c^{Cu}}{m^{Cu}} = \left(\frac{m_c^{Cu}}{m_c}\right)\left(\frac{m_c}{m_r}\right)\left(\frac{m_r}{m}\right)\left(\frac{m}{m^{Cu}}\right) = x_c^{Cu} f_c f_r \frac{1}{x^{Cu}}. \tag{28}$$

Equation 28 is only valid for an open circuit. In reality, the metallurgical recovery is the result of many cycles (e.g., 9) of flotation stages. Because the rougher stage is more widely available, we define the intermediate recovery

$$R_r^{Cu} = x_r^{Cu} f_r \frac{1}{x^{Cu}} \tag{29}$$

and map it to the final recovery with an affine transformation $\alpha + \beta R_r^{Cu}$. These modeling steps lead to the following Bayesian hierarchical model

$$LCT \sim \text{LogitNormal}(\mu, \sigma^2) \tag{30}$$

$$\mu = \alpha + \beta R_r^{Cu} \tag{31}$$

$$\alpha \sim \text{Normal}(a, \sigma^2) \tag{32}$$
$$\beta \sim \text{Normal}(0, \tau^2) \tag{33}$$
$$R_r^{Cu} = x_r^{Cu} f_r \frac{1}{x^{Cu}} \tag{34}$$
$$x^{Cu} \sim \text{LogitNormal}(\mu_x, \sigma_x^2) \tag{35}$$
$$f_r \sim \text{LogitNormal}(\mu_f, \sigma_f^2) \tag{36}$$
$$x_r^{Cu} \sim \text{LogitNormal}(\mu_r, \sigma_r^2) \tag{37}$$
$$\mu_x = \alpha_x + \langle \boldsymbol{\beta}_x, \boldsymbol{x} \rangle \tag{38}$$
$$\mu_f = \alpha_f + \langle \boldsymbol{\beta}_f, \boldsymbol{x} \rangle \tag{39}$$
$$\mu_r = \alpha_r + \langle \boldsymbol{\beta}_r, \boldsymbol{x} \rangle \tag{40}$$
$$\alpha_x \sim \text{Normal}(r_x, \sigma_x^2) \tag{41}$$
$$\alpha_f \sim \text{Normal}(r_f, \sigma_f^2) \tag{42}$$
$$\alpha_r \sim \text{Normal}(r_r, \sigma_r^2) \tag{43}$$
$$\boldsymbol{\beta}_x \sim \text{Normal}(\mathbf{0}, \tau_x^2 \boldsymbol{I}) \tag{44}$$
$$\boldsymbol{\beta}_f \sim \text{Normal}(\mathbf{0}, \tau_f^2 \boldsymbol{I}) \tag{45}$$
$$\boldsymbol{\beta}_r \sim \text{Normal}(\mathbf{0}, \tau_r^2 \boldsymbol{I}). \tag{46}$$

The hyperparameters r_x, r_f, r_r, and a locate the center of the density plots in Fig. 6. They are initialized with empirical averages from the `flotation` table

$$r_x = \frac{1}{n} \sum_{i=1}^{n} x^{Cu\,(i)} \tag{47}$$
$$r_f = \frac{1}{n} \sum_{i=1}^{n} f_r{}^{(i)} \tag{48}$$
$$r_r = \frac{1}{n} \sum_{i=1}^{n} x_r^{Cu\,(i)} \tag{49}$$
$$a = \frac{1}{n} \sum_{i=1}^{n} \text{LCT}^{(i)}. \tag{50}$$

The hyperparameters $\sigma_x, \sigma_f, \sigma_r, \sigma$, and $\tau_x, \tau_f, \tau_r, \tau$ specify the prior uncertainty around these centers. They are initialized with fixed large values. By fixing the hyperparameters of the model, we can sample the prior distribution of x^{Cu}, f_r, x_r^{Cu}, and consequently of LCT as illustrated in Fig. 6.

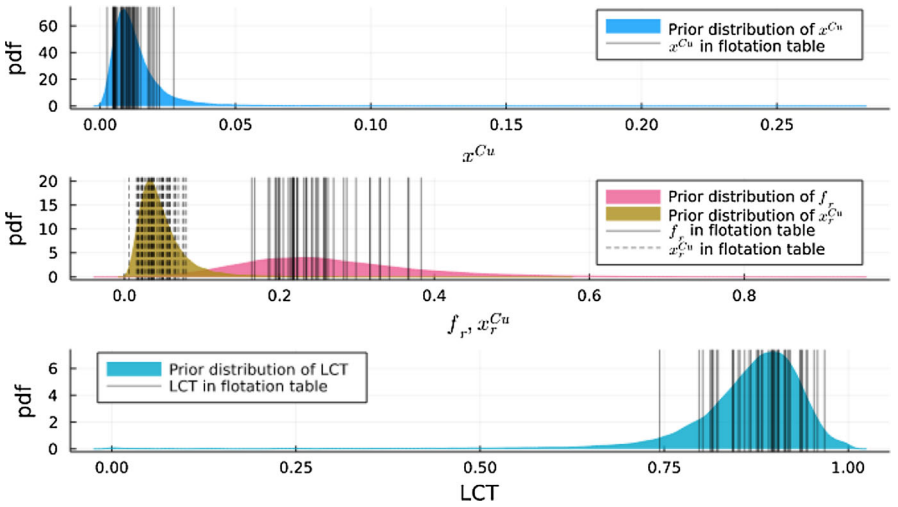


Fig. 6 Prior distribution of x^{Cu}, f_r, x_r^{Cu}, and consequently of LCT. Data from the `flotation` table illustrated with vertical lines

2.3 Inference and Predictions

Having specified the three Bayesian models in Sect. 2.2, we can now proceed with Bayesian inference to obtain the posterior distribution of all the modeled variables given the data from the `comminution` and `flotation` tables. This posterior distribution can then be used to make probabilistic predictions on unseen samples from the `drillholes` table.

In this section, we describe the content of these three tables in more detail, and illustrate predictions of DWT, BWI, and LCT for arbitrarily chosen samples to solidify the concepts presented so far.

2.3.1 Input Tables

The `comminution` table contains data that are used to calibrate the DWT and BWI models. We recall that the DWT index is the product of two parameters that are obtained via least squares and three fractions of fine material known as $t_{10}(h_1)$, $t_{10}(h_2)$, and $t_{10}(h_3)$. Similarly, the BWI is obtained with a standardized formula in terms of $F80$, $P80$, M, and A (see Eq. 12). Therefore, the table must contain all seven of these columns that are directly available from the corresponding laboratory experiments. Besides these columns, the table must also contain the columns with explanatory variables $\boldsymbol{x}$.

The `flotation` table contains data that are used to calibrate the LCT model. We recall that the final metallurgical recovery of copper in the LCT is obtained from the intermediate recovery in the rougher stage, which is in turn obtained from the grade of copper and mass fraction in the rougher concentrate, x_r^{Cu} and f_r, and from the grade of copper in the feed x^{Cu}. Therefore, the table must contain these three variables. Unlike

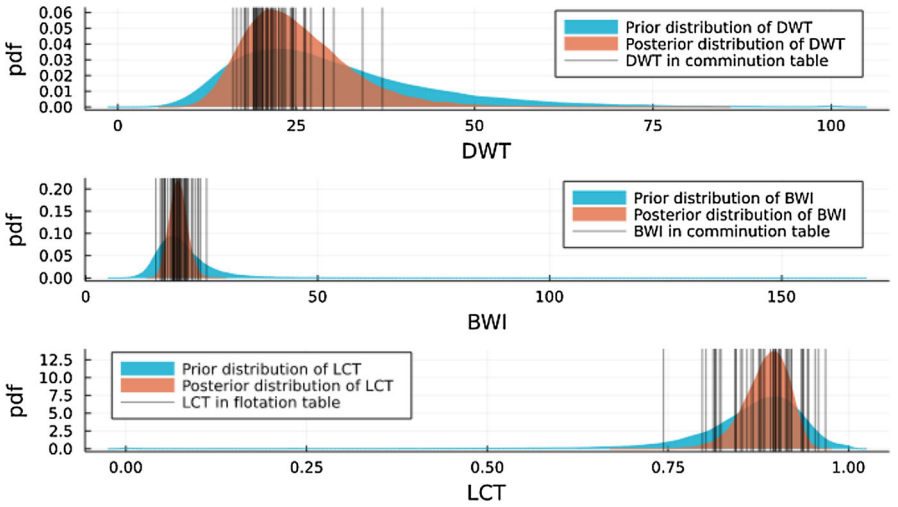


Fig. 7 Prior and posterior distributions of DWT, BWI, and LCT illustrating the relocation of probability to regions where there are a greater number of samples

the other two models, this model also requires a column with the target LCT variable. Besides these four columns, the table must also contain the columns with explanatory variables $\boldsymbol{x}$.

Finally, the `drillholes` table contains the X, Y, and Z columns with the coordinates of samples along the drill holes and the columns with explanatory variables $\boldsymbol{x}$. See the *data availability* section for more details.

2.3.2 Bayesian Inference

Given the models in Sect. 2.2 and the data from the `comminution` and `flotation` tables, we can proceed and perform Bayesian inference with an extension of Hamiltonian Monte Carlo known as the No-U-Turn Sampler (NUTS) (Hoffman and Gelman 2014). The result is a collection of samples from the joint posterior distribution of all variables in the proposed models. In Fig. 7, we illustrate the prior and posterior distributions of DWT, BWI, and LCT. As expected, probabilities are relocated from regions where there are few or no samples to regions where there are a greater number of samples.

We can also visualize the joint posterior distribution of any subset of latent variables, and check that all samples lie inside high-density regions. This is illustrated in Fig. 8 for variables $t_{10}(h_1)$, $t_{10}(h_2)$, and $t_{10}(h_3)$.

Most importantly, it is during Bayesian inference that we learn the joint posterior distribution of coefficients for all explanatory variables. This distribution is used to make predictions on unseen samples, as explained next.

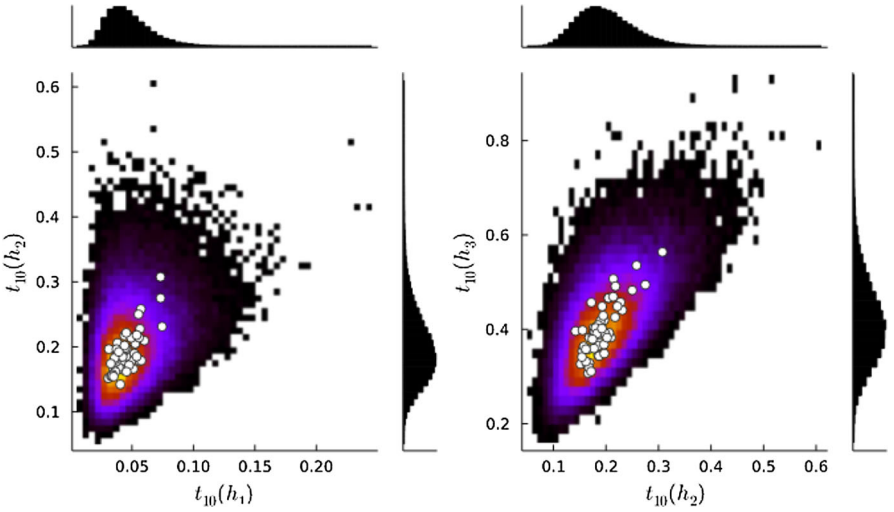


Fig. 8 Joint posterior distribution of $t_{10}(h_1)$, $t_{10}(h_2)$, and $t_{10}(h_3)$. Samples (white dots) in high-density regions (red areas)

2.3.3 Predictions on Drill Holes

For any rock sample with a volume (i.e., support) similar to the volume of the samples in the `comminution` and `flotation` tables, we can use the joint posterior distribution of coefficients to make probabilistic predictions of all other variables in the Bayesian models. We simply evaluate the models forward using the explanatory variables $\boldsymbol{x}$ (e.g., chemical analysis, mineralogy data) of the rock sample and thousands of likely values of coefficients obtained with Bayesian inference. In particular, this procedure generates thousands of likely values of DWT, BWI, and LCT for the rock sample of interest, as illustrated in Fig. 9.

We observe that the posterior mean is not a good estimator of these three geometallurgical variables due to the asymmetry of their distributions, and that a naive approach with Gaussian distributions would have failed to capture this property. We also notice how the specification of the prior distribution is crucial to constrain the values of the variables to physical ranges. Even with a reduced number of samples (≈ 50), we can still obtain useful uncertainty intervals for DWT, BWI, and LCT.

In order to avoid storing thousands of likely values of DWT, BWI, and LCT for each rock sample in the `drillholes` table, we integrate the probability density functions (PDFs) in Fig. 9 into probability mass functions (PMFs) in Fig. 10. This integration consists of counting how many values lie on predefined bins (i.e., histogram). In the case of DWT, we use bins from an industry standard (Chieregati and Delboni Jr 2001): ETA (0,10), EXA (10,20), MTA (20,30), ALT (30,40), MDA (40,50), MED (50,60), MDB (60,70), BAI (70,90), MTB (90,110), ETB (110,120).

The predicted PMFs of DWT, BWI, and LCT for all rock samples in the `drillholes` table are stored in a new `predictions` table along with the X,

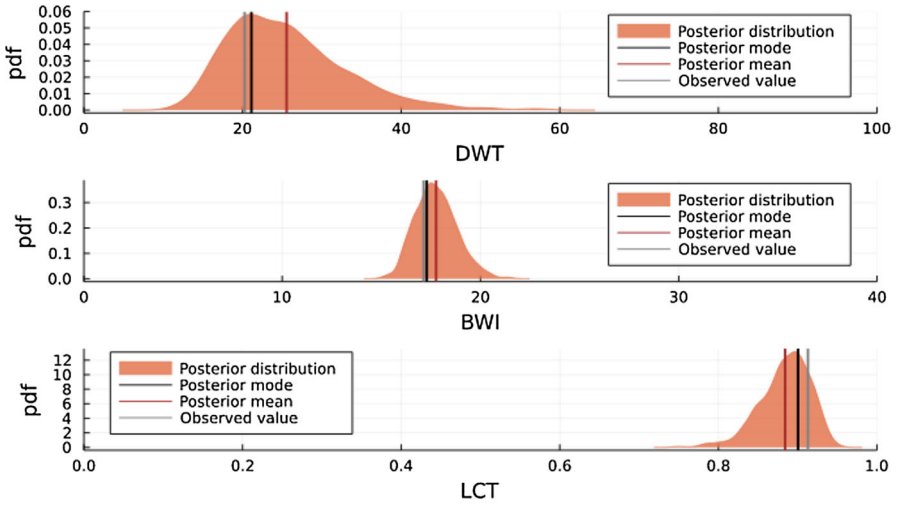


Fig. 9 Probabilistic predictions of DWT, BWI, and LCT for arbitrarily chosen samples from the `comminution` and `flotation` tables. Values observed in the tables lie within the density plots. Posterior mean is not a good estimator of observed values

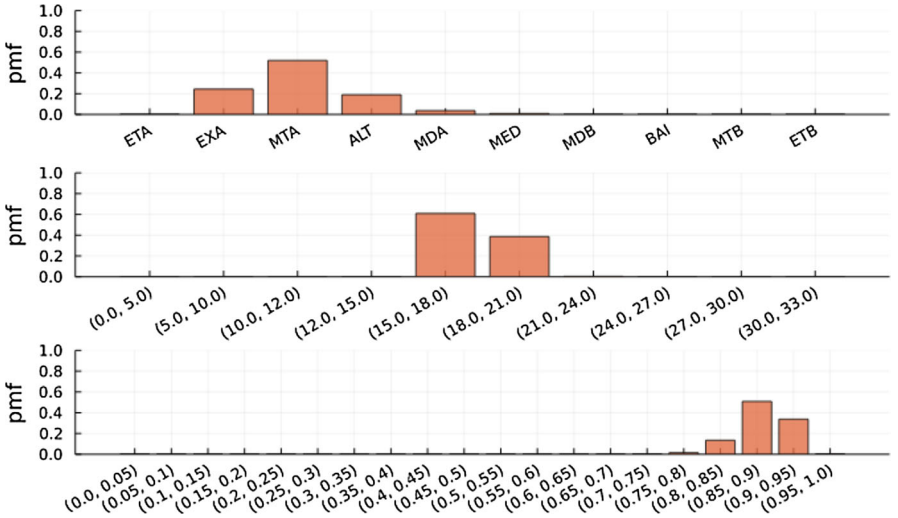


Fig. 10 Corresponding probability mass functions for arbitrarily chosen samples in Fig. 9

Y, and Z coordinates of the samples. This table is the input to the next step of the methodology.

2.4 Hilbert–Kriging

Here we propose the use of Hilbert–Kriging (Menafoglio et al. 2013) for direct interpolation of PDFs (or PMFs) at mining blocks given the `predictions` table. We

briefly review the main concepts behind the framework and explain how an appropriate choice of a Hilbert space can address the issue of non-additivity with geometallurgical variables.

2.4.1 Main Concepts

Consider a collection of k georeferenced objects (e.g., numbers, functions) in a Hilbert space (i.e., vector space with well-defined inner product) $\mathcal{H}$,

$$z(\boldsymbol{s}_1), z(\boldsymbol{s}_2), \ldots, z(\boldsymbol{s}_k) \in \mathcal{H}. \tag{51}$$

where $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k \in D \subset \mathbb{R}^3$ are the locations of the objects. For example, consider vectors with d components (i.e., $\mathcal{H} = \mathbb{R}^d$) and the usual inner product $\langle x, y \rangle = \sum_{i=1}^{d} x_i y_i$, or square-integrable functions (i.e., $\mathcal{H} = L^2$) with the inner product $\langle f, g \rangle = \int f(t)g(t)dt$ (Giraldo et al. 2010). The goal of the Hilbert–Kriging framework is to define an estimator of the object $z(\boldsymbol{s})$ at a new location $\boldsymbol{s} \in D$ as a weighted combination of the available objects

$$\hat{z}(\boldsymbol{s}) = \lambda_1 \cdot z(\boldsymbol{s}_1) + \lambda_2 \cdot z(\boldsymbol{s}_2) + \cdots + \lambda_k \cdot z(\boldsymbol{s}_k) \tag{52}$$

The estimator in Eq. 52 is defined in terms of the scalar multiplication $(\cdot)$ and the vector addition $(+)$ in $\mathcal{H}$. Since the choice of the inner product $(\langle, \rangle)$ induces a norm, a distance, and consequently a notion of variance in $\mathcal{H}$, the universal Kriging system of equations for optimal weights can be generalized to Hilbert spaces via constrained minimization of estimation variance

$$\underset{\lambda_1, \lambda_2, \ldots, \lambda_k}{\text{minimize}} \quad \text{Var}\left(\hat{z}(\boldsymbol{s}) - z(\boldsymbol{s})\right) \tag{53}$$

$$\text{subject to} \quad \mathbb{E}\left[\hat{z}(\boldsymbol{s})\right] = m(\boldsymbol{s}) \tag{54}$$

where $m(\boldsymbol{s}) = \sum_{l=0}^{L} a_l f_l(\boldsymbol{s})$ is the drift, f_l are prespecified monomials in $\boldsymbol{s}$, and a_l are coefficient objects in the Hilbert space.

In practice, the generalization of a modern Kriging implementation to a Hilbert–Kriging implementation consists of two main modifications. First, the coefficient objects must be estimated with an implementation of generalized least squares that supports objects in Hilbert spaces

$$\hat{\boldsymbol{a}} = (\boldsymbol{F}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{F})^{-1} \boldsymbol{F}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{z}, \tag{55}$$

where $\boldsymbol{z} = (z(\boldsymbol{s}_1), z(\boldsymbol{s}_2), \ldots, z(\boldsymbol{s}_k)) \in \mathcal{H}^k$ is the vector with all available objects, $\boldsymbol{\Sigma}$ is the covariance matrix between all objects, $\boldsymbol{F}$ is the monomial matrix for all locations $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_k$, and $\hat{\boldsymbol{a}} = (\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_L) \in \mathcal{H}^{L+1}$ is the vector of estimated coefficients objects (see Sect. 2.4 of the Hilbert–Kriging paper).

Second, the empirical variogram (assuming an intrinsic stationary model) must be estimated in terms of the induced norm ($\|\cdot\|$) in $\mathcal{H}$

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2\,|N(\boldsymbol{h})|} \sum_{(i,j)\in N(\boldsymbol{h})} \left\| z(\boldsymbol{s}_i) - z(\boldsymbol{s}_j) \right\|^2, \tag{56}$$

where $N(\boldsymbol{h}) = \left\{(i,\, j) \colon \boldsymbol{s}_i - \boldsymbol{s}_j \approx \boldsymbol{h}\right\}$ is the usual set of pairs of locations aligned with the lag vector $\boldsymbol{h} \in \mathbb{R}^3$.

Very informally, we say that the choice of scalar multiplication and vector addition determines the geometry of weighted combinations (Eq. 55), and that the choice of inner product determines the structure of geospatial dependence in (Eq. 56) (Menafoglio and Petris 2016). Given that the objects of interest in this work are PDFs (or PMFs) predicted with Bayesian models at drill hole samples, we propose the use of a specific Hilbert space known as the Aitchison space for Hilbert–Kriging interpolation (Menafoglio et al. 2014).

2.4.2 Aitchison Space

In 1986, the statistician J. Aitchison introduced the branch of statistics known today as compositional data analysis to cope with nonlinear constraints on the entries of vector variables (Aitchison 2003). He developed a vector space $\mathcal{A}$ where vectors $\boldsymbol{p} = (p_1, p_2, \ldots, p_m)$ of real entries satisfy

$$p_1, p_2, \ldots, p_m \geq 0 \tag{57}$$

$$\sum_{i=1}^{m} p_i = P. \tag{58}$$

The first constraint, known as the non-negativity constraint, reflects the fact that sometimes entries in a vector only contain relative information (a.k.a. proportions). The second constraint, known as the fixed-sum constraint, exists to guarantee that the initial amount of quantity P is preserved after vector operations. The specific value P is not relevant and is often replaced by $P = 1$ after careful re-normalization.

We note that PMFs satisfy both constraints, and therefore we can leverage the operations of the space $\mathcal{A}$ to interpolate these objects without ever producing invalid probability distributions. The scalar multiplication and vector addition are defined as

$$\lambda \cdot \boldsymbol{p} = \mathcal{C}(p_1^{\lambda}, p_2^{\lambda}, \ldots, p_m^{\lambda}) \tag{59}$$

$$\boldsymbol{p} + \boldsymbol{q} = \mathcal{C}(p_1 q_1, p_2 q_2, \ldots, p_m q_m), \tag{60}$$

with $\lambda \in \mathbb{R}$ a scalar, $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{A}$, PMFs with m bins, and $\mathcal{C}(\boldsymbol{p}) = \frac{\boldsymbol{p}}{\sum_{i=1}^{m} p_i}$ the closure (or re-normalization) operation. As already discussed in the previous section, these operations determine the geometry of weighted combinations. In Fig. 11, we illustrate how two PMFs morph into one another according to these definitions.

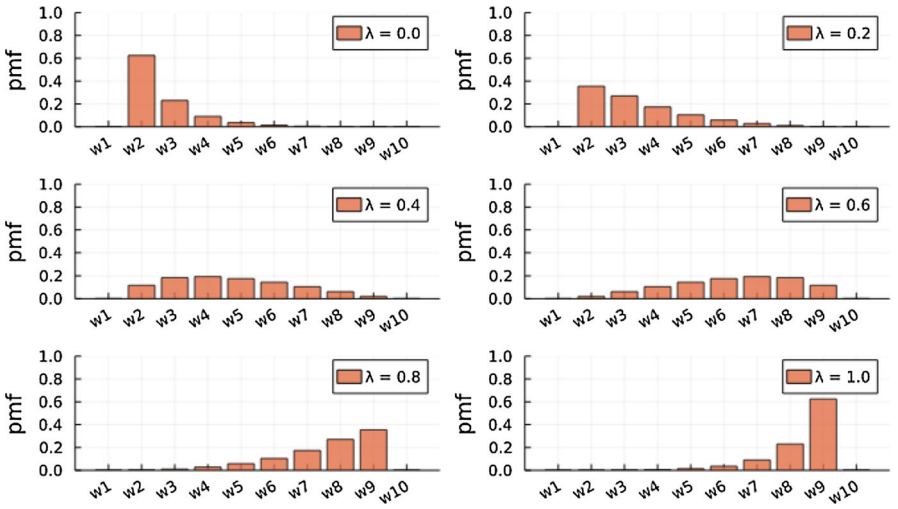


Fig. 11 Scalar multiplication and vector addition in Aitchison space determine how PMFs $\boldsymbol{p}, \boldsymbol{q} \in \mathcal{A}$ (top-left and bottom-right plots) morph into one another as we vary $\lambda \in \mathbb{R}$ in the weighted combination $(1-\lambda)\boldsymbol{p} + \lambda\boldsymbol{q}$

The inner product defined as

$$\langle \boldsymbol{p}, \boldsymbol{q} \rangle = \sum_{i<j} \log \frac{p_i}{p_j} \log \frac{q_i}{q_j} \tag{61}$$

considers ratios (relative information) of entries as opposed to absolute values. It induces a norm $\|\boldsymbol{p}\| = \sqrt{\langle \boldsymbol{p}, \boldsymbol{p} \rangle}$ and a distance $d(\boldsymbol{p}, \boldsymbol{q}) = \|\boldsymbol{p} - \boldsymbol{q}\|$ that can be written as

$$d(\boldsymbol{p}, \boldsymbol{q}) = \sqrt{\sum_{i<j} \left(\log \frac{p_i}{p_j} - \log \frac{q_i}{q_j} \right)^2}. \tag{62}$$

The distance in Eq. 62 determines the structure of geospatial dependence. It is used for empirical variogram estimation in Eq. 56. In Fig. 12, we illustrate how this distance increases as we relocate probability mass across different bins in Fig. 11. It is important to note that these definitions require PMFs with nonzero entries. This is a well-known limitation of the Aitchison space that we overcome with the addition of a very small value to all bins (a.k.a. Laplace smoothing).

Finally, we emphasize that Hilbert–Kriging of PMFs in the Aitchison space enjoys all the features of traditional Kriging, including change of support. In addition, Hilbert–Kriging of PMFs is less prone to non-additivity issues. The scalar multiplication and vector addition in the Aitchison space assure that the asymmetries of the PMFs are preserved and that linear combinations of PMFs are good estimators of the resulting shape of the distribution.

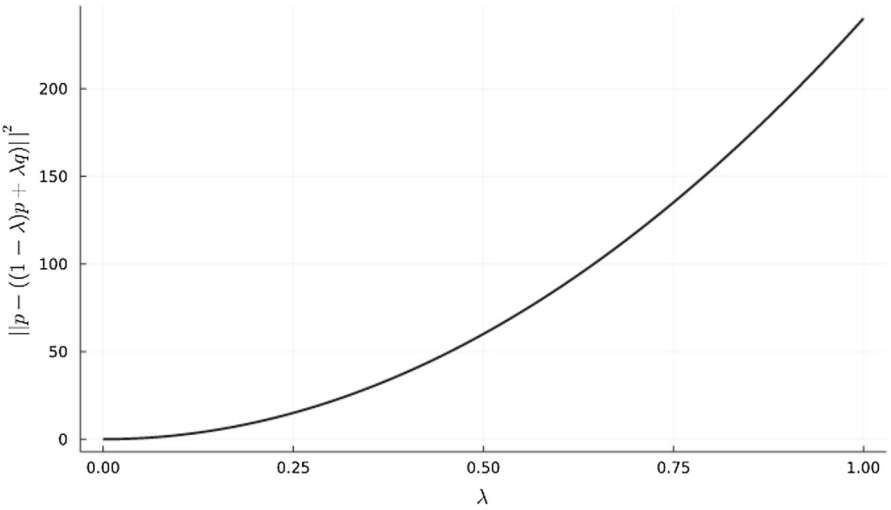


Fig. 12 The inner product in Aitchison space determines distances between PMFsas probability mass is relocated across different bins in Fig. 11

2.4.3 Variography

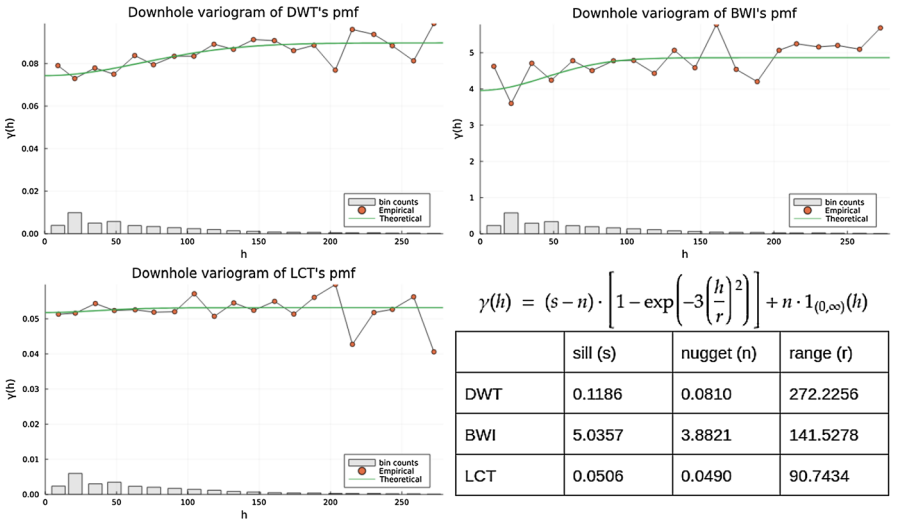
Before we can perform Hilbert–Kriging of PMFs, we first need to model the variogram of these objects. Our goal in this section is to illustrate the shape of such variograms for DWT, BWI, and LCT. Since our intuition for geospatial dependence of PMFs is limited, we do not attempt to interpret these shapes as we usually do with variograms of scalar variables (e.g., ore grades).

Figure 13 illustrates the empirical variograms of the three residual PMFs along the downhole direction after subtracting a drift PMF of zero degree (a.k.a. ordinary Hilbert–Kriging). Because the distance in Eq. 62 is very sensitive to relocation of probability mass (see Fig. 12), these variograms show a high nugget effect. In the same visualization, we plot the corresponding theoretical models fitted with weighted least squares using weights that are proportional to the bin counts.

From the fitted theoretical models in Fig. 13, we observe that the PMFs of DWT and BWI display a visible correlation length (a.k.a. range) for the support of the samples in the `predictions` table ($10m$), whereas the PMF of LCT does not. For simplicity, and because we do not have enough evidence to support an anisotropic theoretical model, we assume that actual variograms of PMFs are not a function of direction, and use these theoretical models fitted along the downhole direction as our omnidirectional models in Hilbert–Kriging.

2.4.4 Geostatistical Estimation

Having modeled the variograms of the three residual PMFs, we can proceed and perform Hilbert–Kriging to estimate the PMFs at all mining blocks of size $30m \times 30m \times 15m$. The change of support is performed as usual by integration of the specified



$$\gamma(h) = (s-n)\cdot\left[1-\exp\left(-3\left(\frac{h}{r}\right)^2\right)\right] + n\cdot 1_{(0,\infty)}(h)$$

	sill (s)	nugget (n)	range (r)
DWT	0.1186	0.0810	272.2256
BWI	5.0357	3.8821	141.5278
LCT	0.0506	0.0490	90.7434

Fig. 13 Empirical variograms of residual PMFs of DWT, BWI, and LCT along the drill hole direction (orange). Theoretical models (green) fitted with weighted least squares using weights that are proportional to the bin counts (gray)

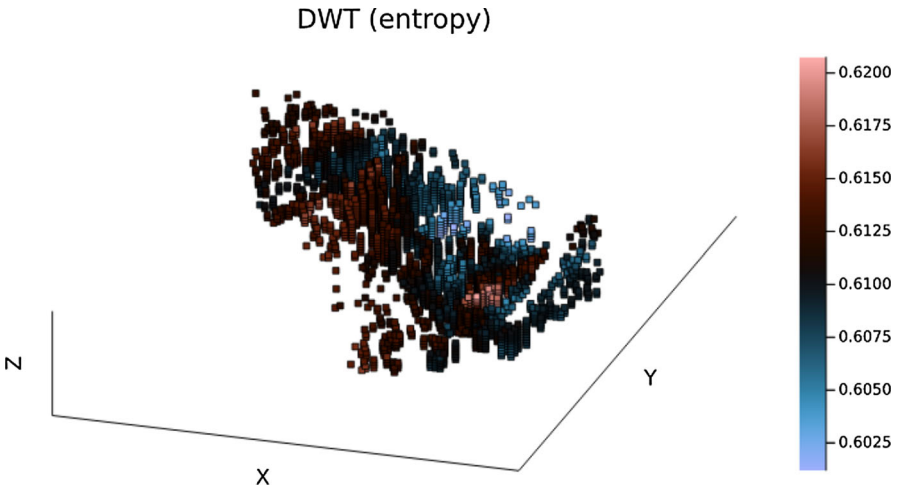


Fig. 14 Geospatial uncertainty of DWT measured by entropy of estimated PMFs

variogram within each block. Let $\boldsymbol{p} = (p_1, p_2, \ldots, p_m)$ be an estimated PMF at a given block. We adopt the mode $\eta(\boldsymbol{p}) = \arg\max_i p_i$ as our final estimate of the corresponding (block support) geometallurgical variable and the entropy $H(\boldsymbol{p}) = -\sum_i p_i \log p_i$ as our measure of uncertainty.

First, we consider the entropy maps of DWT, BWI, and LCT in Figs. 14, 15, and 16. We note that these maps cannot be easily obtained from the geospatial configuration of samples alone in Fig. 2. Unlike the Kriging variance, which is only a function of the geospatial configuration and specified variogram, these maps also encode nonlinear

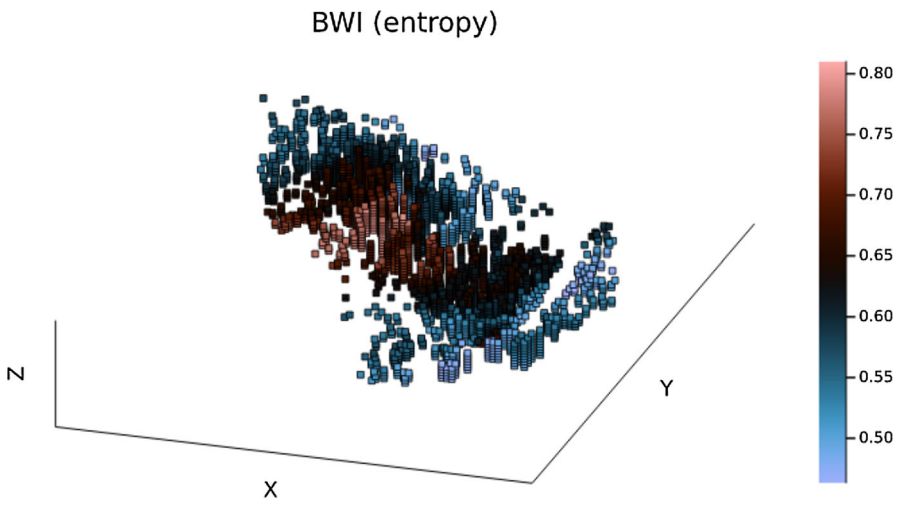


Fig. 15 Geospatial uncertainty of BWI measured by entropy of estimated PMFs

Table 1 Wall times of different steps of the methodology

	Inference (s)	Prediction (s)	Hilbert–Kriging (s)	Total time (min)
DWT	344	177	3	8.7
BWI	83	177	3	4.3
LCT	67	253	3	5.3

Inference performed with $\approx$ 60 samples from the `comminution` and `flotation` tables, prediction performed with 2000 samples from the `drillholes` table, and Hilbert–Kriging performed with $\approx$ 3500 blocks. Wall time per geometallurgical variable is $\approx$ 6 min

relations built into the Bayesian models of Sect. 2.2. In particular, the entropy maps of the two comminution variables—DWT and BWI—are similar, indicating that they may indeed reflect uncertainty about the mechanical competence of the mining block to be extracted.

After considering the geospatial uncertainty, we proceed and look at the mode maps of the three geometallurgical variables. In this specific deposit, the mode maps of DWT and LCT are uninteresting because most predictions of these two variables lie in the same bin interval during the integration of PDFs into PMFs in Sect. 2.3.3. These homogeneous maps are not an issue, since they are just a consequence of the binning choices adopted by the industry. The mode map of BWI is the only heterogeneous map for the specified number of bins. It is shown in Fig. 17.

The block model with estimated PMFs is the final result of the proposed methodology. In Table1, we summarize the wall time of each major step in our implementation of the methodology, which was executed with eight parallel threads in an Intel® Xeon® Platinum 8354H CPU @ 3.10GHz CPU.

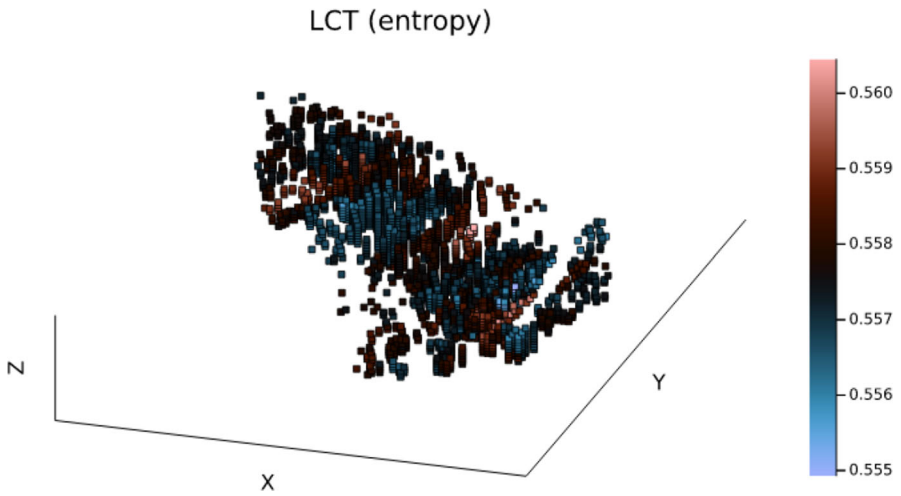


Fig. 16 Geospatial uncertainty of LCT measured by entropy of estimated PMFs

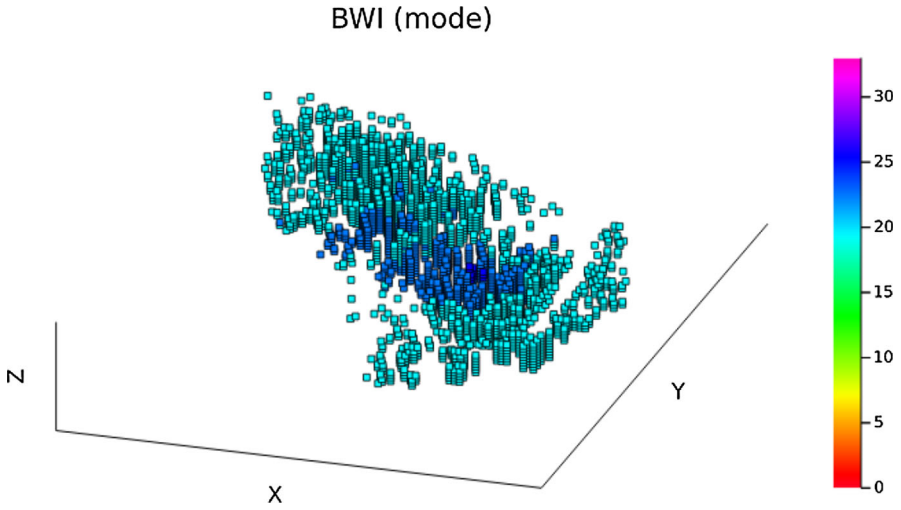


Fig. 17 Most likely interval of BWI measured by mode of estimated PMFs

3 Discussion

As with any other academic work, there are technical challenges that need to be discussed before any serious adoption of the technology by the industry.

In previous sections, we highlighted the major strengths of the proposed methodology, which are:

1. Robust predictions of (non-Gaussian) geometallurgical variables based on domain expertise (e.g., nonlinear relations, prior distributions).

2. Probabilistic predictions from which one can easily obtain uncertainty estimates (e.g., entropy) to assess risk and make strategic decisions.
3. Online predictions that are updated during the development of the mine as soon as new data become available.

We also emphasized that it can generate quick, geospatial, probabilistic estimates of geometallurgical variables without compute-intensive geostatistical simulation (e.g., Gaussian simulation).

In this section, we would like to highlight the weaknesses of the adopted mathematical models and discuss important implementation details that are necessary to make the solution work well in practice. We organize the discussion as a series of issues below.

3.1 Choice of Explanatory Variables

Bayesian models are generally very good at constraining their predictions to acceptable ranges even when the explanatory variables $\boldsymbol{x}$ are not well selected. In this work, we did experiments with transformed chemical compositions (see Sect. 2.2) as our explanatory variables given their widespread availability. However, after a series of experiments with subcompositions, we concluded that these variables may not necessarily be the best predictors of DWT, BWI, or LCT. For instance, we noticed that Bayesian metrics such as the Kullback–Leibler score did not vary much with the addition of new chemical elements to the list of variables. Similar experiments with mineralogical compositions led to the same conclusions.

We believe that the predictive performance of the proposed Bayesian models can be considerably improved with a more thoughtful selection of explanatory variables. In particular, we believe that the resistance of a rock sample to grinding is associated with textural features more than it is associated with chemical or mineralogical compositions. Additionally, extrinsic factors such as operating speeds, circulating volumes, etc. could be taken into account for the prediction of both comminution and flotation variables.

3.2 Covariate Shift

Although the specification of prior distributions gives domain experts control of the likely ranges of all latent variables in Bayesian models, covariate shift of the explanatory variables may still be an issue (Hoffimann et al. 2021). In particular, when the `comminution` and `flotation` tables are too localized in the deposit, it is wise to consider transfer learning methods before making predictions with the `drillholes` table (Pan and Yang 2010; Weiss et al. 2016).

3.3 Compositing Restrictions

The proposed methodology assumes that the samples in the `drillholes` table are composited to lengths that are comparable to the lengths of samples in the

`comminution` and `flotation` tables. If for some reason the comminution and flotation tests are performed with samples of radically varying lengths, it may not be easy to leverage all the data in Bayesian inference.

In practice, laboratory tests are standardized and performed with samples of similar support. Hence, we understand that it is always possible to composite chemical or mineralogical compositions in drill holes to standardized supports in geometallurgy.

3.4 Oversimplification of Physics

To quickly test the methodology with real data from a copper deposit, we opted for oversimplified models of BWI and LCT. The physical processes that are involved in these two tests require dynamical modeling, which was out of the scope of this particular project.

In the case of the BWI, we adopted the Bond formula in Eq. 12, which relies on a series of empirical observations and constants derived decades ago using a specific data set. If the actual deposit shows different behavior, then this behavior can only be incorporated in the model via manual modification of the constants.

Regarding the LCT model, we replaced important aspects of the flotation process by a simple affine map from the rougher stage to the final metallurgical recovery in the locked cycle test. This oversimplification certainly compromises the predictive performance of the model, and requires the measurement of the target LCT variable in the `flotation` table.

3.5 Variable Scaling

Bayesian inference algorithms are sensitive to variable ranges. It is important to scale all variables in a Bayesian model or introduce auxiliary variables to improve the convergence of the associated Markov chains. Fortunately, it is usually possible to scale a variable by a nonzero mean value, or shrink the variable with a nonlinear transformation in the design of new models.

As an example, we introduced the variable $G = P80/F80 \in [0, 1]$ in the BWI model to improve the convergence of the NUTS algorithm with default hyperparameters. Our attempts to work directly with $F80$ and $P80$ values in μm led to low effective sample size and poor mixing of chains.

3.6 Binning Choices

In Sect. 2.3.3, bin intervals were chosen to convert posterior samples into probability mass functions (PMFs). In the case of DWT, these intervals were chosen from an industry standard, whereas in the case of BWI and LCT, the number of bins was chosen to reflect a given level of detail.

It is clear that the choice of bins affects the predicted PMFs. The number of bins must be large enough to effectively approximate the underlying probability density

functions (PDFs) and small enough to make Hilbert–Kriging feasible in conventional computers.

3.7 Laplace Smoothing

The definition of inner product in Eq. 61 assumes nonzero entries in the PMFs, and this is a well-known issue of the Aitchison space. In order to perform variography and Hilbert–Kriging, one must re-normalize the result with the addition of a small threshold value to all entries. The addition of this threshold value can be formulated as Laplace smoothing, which in turn can be interpreted as a form of prior knowledge. Nevertheless, the threshold value is ad hoc, and care must be taken to preserve the original shape of the PMF as much as possible.

3.8 Variogram Interpretation

Unlike variograms of scalar regionalized variables, the variograms used in this methodology are variograms of PMF objects. In this case, the notion of variance relies on the (Fréchet) distance induced by the inner product between PMFs, and consequently it becomes more difficult to associate ranges and sills in Fig. 12 with physical equivalents.

4 Conclusions

In this paper, we addressed the problem of geostatistical interpolation of geometallurgical variables with a novel combination of Bayesian modeling and Hilbert–Kriging. The proposed methodology produces quick ($\approx$ 6 min), geospatial, probabilistic estimates of non-Gaussian variables without ad hoc transformations and compute-intensive geostatistical simulation.

We applied the methodology to a real copper deposit to illustrate that it can work in practice at industrial scale. By considering three geometallurgical variables from comminution and flotation tests, we showed that the proposed approach honors the shape of the posterior (non-Gaussian) distribution at each mining block, including their characteristic asymmetry. Additionally, all these probabilistic predictions presented satisfactory geospatial continuity according to simple visualizations of mode and entropy maps.

One of the main practical challenges encountered during the application of the methodology was the interpretation of variograms of probability mass functions. In this work, we assumed that the empirical variograms along the downhole direction were a good approximation of an omnidirectional model. Further experiments with different data sets are needed to assess the extent to which this assumption is valid, and to gain more intuition about this notion of geospatial continuity.

Finally, we believe that this work proves the concept that more sophisticated mathematical models are well suited to address major challenges in geometallurgical modeling. Future work could consider new Bayesian models inspired by physics,

new data sets from different types of deposits, and more quantitative assessments designed in partnership with the industry.

Acknowledgements The authors acknowledge the leadership of the IMPA–Vale partnership, a research collaboration between the Instituto de Matemática Pura e Aplicada and Vale S.A.

Data availability Obfuscated versions of the `comminution`, `flotation`, and `drillholes` tables are publicly available at: https://zenodo.org/record/6587598 (Hoffimann et al. 2022). The data obfuscation procedure does not affect any of the presented results.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Software citations This project was developed with state-of-the-art open-source libraries and the Julia programming language (Bezanson et al. 2017). The proposed Bayesian models were written with Turing.jl, a general-purpose probabilistic programming language for robust, efficient Bayesian inference and decision making. (Ge et al. 2018). The Hilbert–Kriging method is available in GeoStats.jl, an extensible framework for high-performance geostatistics (Hoffimann 2018).

Authors Contributions We adopt the CRediT taxonomy. Júlio Hoffimann: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Visualization, Supervision, Project administration. José Augusto: Methodology, Software, Formal analysis, Investigation, Data Curation. Lucas Resende: Methodology, Software, Formal analysis, Investigation, Data Curation. Marlon Mathias: Software, Formal analysis, Data Curation. Douglas Mazzinghy: Conceptualization, Validation, Resources, Writing - Review & Editing. Matheus Bianchetti, Mônica Mendes, Thiago Souza, Vitor Andrade, Tarcísio Domingues: Validation, Data Curation. Wesley Silva, Ruberlan Silva, Danielly Couto: Data Curation. Elisabeth Fonseca: Conceptualization, Validation, Resources, Data Curation, Project administration. Keila Gonçalves: Funding acquisition.

Appendix A DWT Is Not Linear

Given the objective function $J_i(A,b) = \left\|\boldsymbol{t}^{(i)} - A\left(\mathbf{1} - e^{-b\boldsymbol{E}}\right)\right\|_2^2$, we show that the optimal parameters $A^{(i)}, b^{(i)} = \arg\min_{A,b>0} J_i(A,b)$ cannot be both linear functions of a common set of explanatory variables $\boldsymbol{x}$.

First, we note that the derivatives must vanish at the optimal solution

$$\frac{d}{dA}J_i = -2\left(\boldsymbol{t}^{(i)} - A\left(\mathbf{1} - e^{-b\boldsymbol{E}}\right)\right)^\top\left(\mathbf{1} - e^{-b\boldsymbol{E}}\right) = 0 \tag{63}$$

$$\frac{d}{db}J_i = -2A\left(\boldsymbol{t}^{(i)} - A\left(\mathbf{1} - e^{-b\boldsymbol{E}}\right)\right)^\top\left(\boldsymbol{E}\circ e^{-b\boldsymbol{E}}\right) = 0 \tag{64}$$

where $\circ$ is the Hadamard (or entry-wise) product. From the first equation, we obtain the following expression for $A^{(i)}$ in terms of $b^{(i)}$

$$A^{(i)} = \frac{\langle\boldsymbol{t}^{(i)}, \mathbf{1} - e^{-b^{(i)}\boldsymbol{E}}\rangle}{||\mathbf{1} - e^{-b^{(i)}\boldsymbol{E}}||_2^2}. \tag{65}$$

From the second equation, we obtain the following implicit relation for $b^{(i)}$ after substituting the expression obtained above for $A^{(i)}$,

$$\left\langle \boldsymbol{t}^{(i)} - \frac{\langle \boldsymbol{t}^{(i)}, \mathbf{1} - e^{-b^{(i)}\boldsymbol{E}} \rangle}{||\mathbf{1} - e^{-b^{(i)}\boldsymbol{E}}||_2^2}(\mathbf{1} - e^{-b^{(i)}\boldsymbol{E}}), \boldsymbol{E} \circ e^{-b^{(i)}\boldsymbol{E}} \right\rangle = 0. \tag{66}$$

Denoting by

$$proj_{\left(\mathbf{1}-e^{-b^{(i)}\boldsymbol{E}}\right)}\boldsymbol{t}^{(i)} = \frac{\langle \boldsymbol{t}^{(i)}, \mathbf{1} - e^{-b^{(i)}\boldsymbol{E}} \rangle}{||\mathbf{1} - e^{-b^{(i)}\boldsymbol{E}}||_2^2}(\mathbf{1} - e^{-b^{(i)}\boldsymbol{E}}), \tag{67}$$

we can rewrite the implicit relation as

$$\left\langle \boldsymbol{t}^{(i)} - proj_{\left(\mathbf{1}-e^{-b^{(i)}\boldsymbol{E}}\right)}\boldsymbol{t}^{(i)}, \boldsymbol{E} \circ e^{-b^{(i)}\boldsymbol{E}} \right\rangle = 0. \tag{68}$$

We now know that the pair $(A^{(i)}, b^{(i)})$ of optimal parameters must satisfy Eqs. 65 and 68. It is clear that if $b^{(i)}$ is a linear function of $\boldsymbol{x}$, then $A^{(i)}$ cannot be a linear function of $\boldsymbol{x}$, and vice versa. For similar reason, the product $\text{DWT}^{(i)} = A^{(i)} \times b^{(i)}$ cannot be a linear function of $\boldsymbol{x}$.

References

Adeli A, Dowd P, Emery X, Xu C (2021) Using cokriging to predict metal recovery accounting for non-additivity and preferential sampling designs. Miner Eng 170:106923

Agar G (2000) Calculation of locked cycle flotation test results. Miner Eng 13:1533–1542

Aitchison J (2003) The statistical analysis of compositional data. Blackburn Press, Caldwell

Barnett RM, Deutsch CV (2012) Practical implementation of non-linear transforms for modeling geometallurgical variables. Springer, Dordrecht

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: a fresh approach to numerical computing. SIAM Rev 59(1):65–98

Boisvert J, Rossi M, Ehrig K, Deutsch C (2013) Geometallurgical modeling at Olympic Dam Mine, South Australia. Math Geosci 45

Bond F (1961) Crushing and grinding calculations parts 1 and 2. Br Chem Eng 6(378–385):543–548

Campos PHA, Costa JFCL, Koppe VC, Bassani MAA (2021) Geometallurgy-oriented mine scheduling considering volume support and non-additivity. Min Technol 1–11

Carrasco P, Chilès JP, Séguret S (2008) Additivity, metallurgical recovery, and grade. VIII International Geostatistics Congress, GEOSTATS 2008:1188

Chieregati A C, Delboni Jr H (2001) Nova metodologia de caracterização de minérios aplicada a projetos de moinhos ag/sag. In VI SHMMT/XVIII ENTMME

Davidson-Pilon C (2015) Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference. Addison-Wesley Professional, 1st edition

Deutsch JL, Palmer K, Deutsch CV, Szymanski J, Etsell TH (2015) Spatial modeling of geometallurgical properties: techniques and a case study. Nat Resour Res 25(2):161–181

Ge H, Xu K, Ghahramani Z (2018) Turing: a language for flexible probabilistic inference. International conference on artificial intelligence and statistics, AISTATS 2018, 9–11 April 2018. Playa Blanca, Lanzarote, Canary Islands, Spain, pp 1682–1690

Gelman A (2014) Bayesian data analysis. CRC Press, Boca Raton

Giraldo R, Delicado P, Mateu J (2010) Ordinary kriging for function-valued spatial data. Environ Ecol Stat 18(3):411–426

Hoffimann J (2018) Geostats.jl - high-performance geostatistics in Julia. J Open Source Software 3(24):692
Hoffimann J, Augusto J, Resende L, Mathias M, Mazzinghy D, Bianchetti M, Mendes M, Souza T, Andrade V, Domingues T, Silva W, Silva R, Couto D, Fonseca E, Gonçalves K (2022) Geomet dataset
Hoffimann J, Zortea M, de Carvalho B, Zadrozny B (2021) Geostatistical learning: Challenges and opportunities. Front Appl Math Stat 7
Hoffman MD, Gelman A (2014) The no-u-turn sampler: adaptively setting path lengths in Hamiltonian monte Carlo. J Mach Learn Res 15:1593–1623
Journel AG (2003) Mining geostatistics. Blackburn Press, Caldwell
Menafoglio A, Guadagnini A, Secchi P (2014) A kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. Stoch Env Res Risk Assess 28(7):1835–1851
Menafoglio A, Petris G (2016) Kriging for Hilbert-space valued random fields: the operatorial point of view. J Multivar Anal 146:84–94
Menafoglio A, Secchi P, Rosa M D (2013) A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. Electron J Stat 7(none):2209–2240
Napier-Munn T, Morrell S, Morrison R, Kojovic T (1999) Mineral comminution circuits their operation and Optimisation. JKMRC Monograph Ser Mining Mineral Process 2
Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359
Rowland JCA (1975) The tools of power: how to evaluate grinding mill performance using the Bond work index to measure grinding efficiency. In AIME Annual Meeting, Arizona
Tavares LM, Kallemback RD (2013) Grindability of binary ore blends in ball mills. Min Eng 41:115–120
Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J Big Data 3(1)