A self-adaptive particle-tracking methodology for minerals processing

Pereira, L.; Frenzel, M.; Khodadadzadeh, M.; Tolosana Delgado, R.; Gutzmer, J.;

Originally published:

August 2020

Journal of Cleaner Production 279(2021), 123711

DOI: https://doi.org/10.1016/j.jclepro.2020.123711

Perma-Link to Publication Repository of HZDR:

https://www.hzdr.de/publications/Publ-30727

A self-adaptive particle-tracking method for minerals processing

Lucas Pereira, Max Frenzel, Mahdi Khodadadzadeh, Raimon Tolosana-Delgado, Jens Gutzmer

Word count (abstract-conclusions): 6624

Word count (total): 8625

Abstract

Modern analytical techniques used in the minerals processing industry can provide detailed characterization data at the particle level. However, process models that make full use of this information are currently not available, limiting the usefulness of these extensive datasets. This contribution addresses this issue. It presents a novel particle-based approach for process modelling capable of dealing with complete particle datasets and operating without human input. The method provides a probabilistic description for the behavior of individual particles in a given mineral processing unit, based on all measurable particle properties. It is applicable to any separation process that does not modify the physical dimensions of particles, i.e. it does not cover comminution.

The method comprises a regularized logistic regression model with a probability adjustment step to accommodate geological variability. Even though this method supports any type of particle-level characterization data, its potential is illustrated here using data obtained by scanning electron microscope-based image analysis. Constructed cases demonstrate the efficiency of the method in recreating characteristic recovery trends for magnetic separation, hydrocyclone, and flotation units. In addition, the method was used successfully to reconstruct a real processing plant with three flotation and one magnetic separation circuits. Predicted results of compositions for all the intermediate and product streams correspond well with the results reported from the plant itself. The predicted masses of the products are much affected by the quality of sampling and still require improvement. The case study illustrates that the method proposed here provides a powerful tool to understand and optimize mineral separation processes – and thus increase the resource and energy efficiency of mining operations.

Keywords: Particle-tracking; geometallurgy; mineral processing modelling; flotation; resource efficiency; automated mineralogy

1 Introduction

Currently, most mineral beneficiation plants are controlled and optimized using bulk material characterization data (Hodouin, 2011). This is despite the fact that modern analytical methods can yield detailed mineralogical and geometrical information for individual particles in a raw material stream, which can in turn be used to understand the behavior of the material in different unit operations. Provided that the feed material can be suitably characterized and that robust models for the unit operations are available, it should be possible to use particle-based data to predict the efficiency and

selectivity of each unit. It is a major goal of the circular economy (Kirchherr et al., 2017) and sustainable production practices to optimize resource efficiency, e.g., by providing a clear understanding of particle recoverability in minerals processing circuits. This pertains to conventional minerals processing operations as well as tailings dam reclamation (Büttner et al., 2018) or by-product recovery (Frenzel et al., 2019; Pereira et al., 2019).

In this article, we present a new method for particle-based process modelling that quantifies the probability of each particle to be recovered into a concentrate stream according to its properties. No human input is required, and the method is self-adaptive to any minerals processing unit that does not modify the physical dimensions of the particles.

1.1 Background

Minerals processing machinery operates by exploiting the physical and chemical properties of ore and gangue minerals to achieve efficient separation (Wills and Finch, 2015). For example, gravity separation devices (e.g. hydrocyclones) create conditions in which particles of different sizes and densities follow distinct trajectories. Flotation, a more complex technique, exploits differences in surface chemistry of the different mineral groups.

All particle separation processes are best described in a probabilistic manner using the long-established concept of recovery probability as a function of particle characteristics (Jowett, 1986; Tromp, 1950, 1937). A practical example is the hydrocyclone partition curve (King et al., 2012; Wills and Finch, 2015), which gives the probability of a particle to enter the underflow as a function of its density and size.

Probabilities are also a useful measure to compare the effect of different equipment set-ups on the recoverability of a particle, e.g., when optimizing a processing unit for a specific ore type (Jowett, 1986; Schach et al., 2019; Tromp, 1937). However, the extent to which particle-based process description was possible in the past was limited by both the available computing power as well as a lack of suitable characterization techniques. Thus, most available modelling tools can only process particle property distributions rather than whole particle property datasets (King et al., 2012). Recently, the use of advanced Machine Learning (ML) techniques has been explored to enhance numerical models and improve predictive power (McCoy and Auret, 2019). The main appeal of ML is its capability to handle the unprecedented amount of complex data produced by modern analytical techniques (e.g. Ilisei et al., 2019; Khodadadzadeh et al., 2014; Maxwell et al., 2018).

Several analytical methods are now available that provide detailed characterization data at the single-particle level. The most important ones are scanning electron microscope (SEM)-based image analysis (Fandrich et al., 2007; Sandmann, 2015), X-ray computed tomography (Godinho et al., 2019; Ketcham and Carlson, 2001) and optical microscopy (Petruk, 2000). SEM-based image analysis, arguably the most relevant analytical technique in minerals processing research in the last two decades, quantifies both the composition and the dimension of particles in 2D sections. The method is well suited to characterize particle streams with P_{50} grain sizes lower than ~1 mm and greater than ~5 µm. Other methods may be more suitable beyond this particle size range.

A particle-based approach to process prediction, called particle tracking, was first introduced by Lamberg and Vianna (2007). It uses the mineralogy, surface composition and particle size data from SEM-based image analysis systems. Five steps of data binning, treatment and smoothing, one step of data reduction, and one step of chemical and mineralogical data integration are necessary before predictions can be made. These complex data-manipulation steps force an internal consistency on the data, which may or may not exist. In addition, the method requires significant human input (e.g. for elemental to mineral conversion) that might lead to biased estimates. While it was successful in predicting the outcome of a specific sulfide flotation test (Lamberg and Vianna, 2007), limitations due to unbalanced particle bins quickly appeared when applied to a case of iron ore separation with the Davis Tube (Cárdenas, 2017).

At the same time, Pascoe et al. (2007) introduced a similar approach to calculate particle recovery from SEM-based image analysis data. In this method, particles are binned according to density and size. Recovery is then expressed by the percentage of particles in each bin reporting to the concentrate fraction. This approach was used to predict particle behavior in a gravity separation unit. Its main drawbacks are the limitation to two variables and the decrease in data resolution due to binning.

Recently, Hannula et al. (2018) proposed a modification of these previous particle tracking methods. The authors introduced a dynamic binning strategy in which the user selects the particle properties to be used for the simulation of a specific process, and the variable distribution defines the binning classes – an improvement towards the statistical representation of particles. In addition, the authors applied neural networks to generalize the predictive model and provide continuous recovery probabilities for each bin.

Finally, Schach et al. (2019) created a method that directly links particle-tracking to partition curves. For this, the authors used kernel density estimates to quantify continuous probabilities over the complete 2D sample space defined by particle density and size. While the method can in principle be extended to higher dimensions, the total number of variables is limited by computational complexity and data density to less than ~10. A pre-selection of variables is therefore still required.

To our knowledge, no user-independent particle-tracking method that does not require any data reduction step (neither variable selection nor particle binning) has been developed so far. When experts dictate the variables to be used in such methods, they apply knowledge gained in times when detailed characterization data was not available. This strategy may not only add human bias to the results, but also hinders the exploration of the influence of “new” variables such as particle mineralogy and surface composition in separation processes.

To fill this gap, this contribution presents a new particle-tracking method that makes full use of complex particle datasets with minimum human input required, and that is self-adaptive to different processing unit types. The method delivers quantitative measures for particle behavior in the form of probabilities. It can be extended to model the effect of variations in machine operating conditions and thus has the potential to become a powerful tool for process optimization.

2 Methodology

This section gives a brief overview of the assumptions made in the design of the new method, the structure of particle datasets and the data treatment steps applied, followed by a complete description of the algorithm. Even though the new particle-tracking method can be applied to particle datasets acquired by different characterization techniques, it is demonstrated in this paper using SEM-based image analysis data.

2.1 Assumptions and limitations

The proposed method is aimed at identifying patterns and creating predictive models for different process units by comparing the properties of particles present in each of the output streams. Several assumptions are implicit in this approach:

1) Sample representativity. Analyzed samples are assumed to be representative of the process streams they were taken from.
2) Particle preservation. The separation process is assumed not to modify physical particle properties (e.g. size), with the exception that modifications to surface chemistry, e.g. during froth flotation, are allowed.
3) Analytical representativity. Measured particle properties are assumed to be representative of the true properties of the particles.

While assumption 1) is generally fulfilled as long as best-practice sampling (Lotter, 2011; Wills and Finch, 2015) and sample preparation (e.g. Heinig et al., 2015) procedures are observed, assumptions 2) and 3) provide some limitations towards the use of the new method. Specifically, assumption 2) means that the method cannot be applied to any process involving particle modification, particularly comminution.

Assumption 3) refers to the stereological bias associated to analytical techniques that measure the 2D sections of particles rather than true 3D characteristics, such as SEM-based image analysis. While particle properties such as particle size or mineral liberation measured under these conditions generally still correlate strongly with true characteristics, there is a limit to the accuracy with which prediction is possible (cf. Miller et al., 2009; Ueda et al., 2018). We nevertheless use such data here, since SEM-based image analysis is currently the most suitable technique for the analysis of the mineralogical complexity and particle size ranges encountered in the processing of complex ores (cf. Sandmann, 2015). To date, no other analytical method can provide datasets with similar resolution. We further note that stereological bias influences all samples measured under the same conditions in a similar way, such that measured particle properties should be comparable within a sample set. This should tend to limit the detrimental effects of stereological bias.

Finally, it is worth noting that the nature of the method is such that predictions are only possible within the parameter space covered by the samples included in the training data for the models. If the models are extrapolated to completely new material types, they are not expected to provide reliable results.

2.2 Data structure and required pre-treatment

Particle datasets typically contain information on each particle's modal composition (in units of volume or weight), surface composition (area), and different measurements of size (e.g. equivalent circle diameter, maximum Feret diameter) and shape (e.g. aspect

ratio, solidity). In addition, these variables can be combined to yield further information such as particle density and chemical composition. The set of minerals in each ore is unique and defines the total number of variables. Sandmann (2015) presents more details about SEM-based image analysis datasets.

Before collating this multitude of data types together, most of them require pre-treatment or transformation. Below, the treatments applied in the present case are described in detail. Note, however, that particle datasets acquired by analytical methods other than SEM-based image analysis, e.g. hyperspectral-imaging or X-ray computed tomography, may require different pre-treatments. This is particularly true where they contain data types not included in the list below such as spectral absorbance features for hyperspectral imaging, or grey-value data for X-ray computed tomography.

2.2.1 Particle bulk and surface compositions

In SEM-based image analysis data, the bulk mineralogical make-up of a particle is generally described via the volumes or masses contributed by different minerals, and particle surface composition is described by mineral area. To remove the strong correlation between these quantities and particle size, we normalized all data referring to particle compositions to a sum of 100 %. This was done separately for the bulk and surface compositions of each particle. The subsequent log-ratio transformation that is usually recommended for compositional data was not applied in our method (cf. van den Boogaart and Tolosana-Delgado, 2013). The main reason for this is that particle composition data is usually strongly zero-inflated (Barry and Welsh, 2002), and most of the zeroes are true zeroes, such that imputation is not a suitable option (cf. van den Boogaart and Tolosana-Delgado, 2013). Furthermore, mineral groupings were avoided to minimize human bias.

2.2.2 Particle dimension properties

2.2.2.1 Size data

Different measurements of size are available for 2D image analysis, each of them with advantages and disadvantages depending on particle shape. Equivalent circle diameter (*ECD)* is one of the most widely used measurements of size in SEM-based image analysis (Sandmann, 2015), with particular advantages for quasi-spherical particles. Since ECD is measured exclusively on a positive scale, ECD values are expected to follow a log-normal distribution (Gaddum, 1945), which has long been used by mineral processing engineers to describe particle size distributions (King et al., 2012; Wills and Finch, 2015). Thus, a log-transformation was applied to ECD for data normalization.

Furthermore, it was necessary to include the square of the log-transformed ECD so that the fitted regularized logistic regression model would be capable of reconstructing the roughly parabolic shape of the dependence of flotation recovery on log-particle size (King et al., 2012; Wills and Finch, 2015). This is illustrated with a constructed test case in section 3.1.

2.2.2.2 Shape data

To compensate for the use of a size variable that can be biased according to a particle's shape, two shape parameters are also added to the dataset. Solidity, *S*, is the measure

of the overall concavity of a particle. It is defined as the particle area divided by the particle convex hull area: $S = A/A_c$ (Olson, 2013). Aspect ratio, *AR*, relates to the overall shape of a particle and can be calculated by dividing the particle minimum Feret diameter by its maximum Feret diameter: $AR = x_{Fmin}/x_{Fmax}$ (Olson, 2013). To ensure that the correction of ECD by these shape parameters is easily possible within the model, they were also log-transformed.

2.2.2.3 Unique dimension influence on each mineral

In order to capture a mineral-specific effect of particle dimension properties on the recoverability of a particle, a categorical variable that indicates the main mineral in mass of each particle is required. An interaction between continuous and categorical variables in the form of an analysis of covariance model (Keppel and Wickens, 2004) is thus added to the particle dataset.

2.3 Algorithm

To model a processing unit, particle data from its output streams is collected, treated, and merged to form a training dataset. For this merging, the same fixed number of particles is sampled with replacement from each stream, resulting in equal numeric proportions of particles from all streams. A class label is then added to each particle to indicate its source stream.

A least absolute shrinkage and selection operator (lasso)-regularized logistic regression model (LR) is then trained as a classifier (Hastie et al., 2015). Logistic regression is a robust statistical method commonly used to model the probabilities of individual observations to belong to specific classes (Cox, 1958). Even though logistic regression is mathematically able to deal with high-dimensional datasets, it commonly fails in the assignment of meaningful coefficients to each variable. In such cases, regularization is required and the lasso-regularized regression is commonly applied for the task. This technique uses a penalty factor to set the coefficients for non-significant variables to zero and estimate more reliable coefficients for the remaining predictors. The best penalty factor is identified for each case through cross-validation (model accuracy maximization). As a result, the LR is sparser and has a lower tendency to overfit the data (Hastie et al., 2015). In its original form, it can only deal with binomial classes (e.g. “concentrate” vs. “tailings”). However, an extension to additional outcome classes exists with lasso-regularized multinomial logistic regression (Khodadadzadeh et al., 2014). This also makes the method suitable for the modelling of separation processes with multiple outcome streams such as shaking tables.

In summary, the logistic regression consists of the use of a logistic function to model a binary dependent variable. It works by fitting a linear combination of variables to the log-odds of the binary dependent variable, and the conversion of log-odds to probabilities can be done using the logistic function (Cox, 1958).

The fitted coefficients of the LR model can be interpreted as the impacts that the different variables have on the probabilities of each particle to behave in a certain way in the process (Hastie et al., 2015). Since the data treatment step does not guarantee similar scales among variables, their coefficients do not present comparable scales and should be interpreted with particular care (Hastie et al., 2015). The sign of each coefficient is related to the definition of the probabilities used to describe the process. For instance, if the probability of a particle to report to the concentrate is predicted,

positive coefficients indicate variables that have a positive influence on particles to report to concentrate.

It is important to mention here that other classifiers such as the random forest (Breiman, 2001), linear discriminant analysis (Ripley, 1996), and XG-Boost (Chen and Guestrin, 2016) could in principle be used in our proposed system. However, LR is the best option in this case given the data complexity and need to obtain reliable probability estimates (De Vasconcelos et al., 2001; Khodadadzadeh et al., 2014; Wilks, 2009).

After training, the resultant LR model can be applied to a particle dataset representing the actual feed material of the corresponding processing unit to predict the likelihood of the possible processing outcomes for each particle. However, the influence of the relative abundances of different particle classes in the feed on the predicted probabilities needs to be taken into account. Following Bayes' theorem, the posterior probability of a particle to report to a stream can be obtained given its likelihood and prior probability to belong to that stream.

In the case of particle tracking, the prior probability of a particle to report to a given process stream can be taken as the overall frequency at which particles report to that stream. This is closely related, but not equal, to the system's mass balance.

While prior probabilities can be easily estimated from the mass balance for the training data, it is extremely unlikely that a new feed material will have the same proportions of concentrate- and tailings-like particles. In fact, the opposite would be expected given the natural variability of ore deposits. In such cases, we suggest that the method developed by Saerens et al. (2002) be used to estimate the prior probabilities for every new feed based on its composition – and accordingly correct each particle's posterior probability.

However, if the prior probabilities of a new sample are the same as in the training set, this prior estimation procedure is not advisable since it may lead to erroneous estimations (Saerens et al., 2002). The authors provided a test to define whether the correction should be performed or not.

For process simulation, the probabilities obtained from the LR model can be used to assign feed particles to each of the output classes. In accordance with accepted conceptual models for minerals processing operations (Tromp, 1937, 1950), this assignment is random, using the modelled probabilities to assign a class label to each particle. Following the assignment of all particle labels, the overall composition of the output streams can then be characterized in terms of their mass, modal mineralogy, particle size distribution etc.

The random nature of particle class assignments introduces uncertainties in the overall result. Therefore, bootstrapping is applied to determine the most probable outcome of this probabilistic process, as well as the relevant confidence intervals. The overall uncertainty reflected in these confidence intervals will be chiefly related to sample size and, in the present case, some of the random deviations of 2D particle cross-sections from their true 3D properties. Systematic errors introduced during sampling or analysis will not be captured by these uncertainty estimates.

The bootstrapping routine (Henderson, 2005) is implemented as follows: the feed dataset is repeatedly sub-sampled with replacement (same number of particles) to simulate the natural variability of the ore. The posterior probability calculation, class assignments, and overall characterization of each stream are then repeated, yielding a range of aggregate values for each property (e.g. modal mineralogy). This is done 1000 times to get reliable estimates of median values and confidence intervals. Figure 1 summarizes this workflow.

3 Demonstration

The functionality of the method is demonstrated in this section using several artificial and real cases. All calculations were done in R (R Core Team, 2017) using package *"glmnet"* for the lasso-regularized logistic regression (Friedman et al., 2010), and package *"ggplot"* for the plots of results (Wickham, 2016).

3.1 Artificial test cases

To test the ability of our method to correctly predict the probabilities of individual particles to report to a given process stream, several artificial cases were constructed. This was necessary since particle-specific probabilities cannot be obtained independently for any real dataset. We therefore used part of a treated particle dataset (100,000 particles; 10 predictors) in combination with quantitative models of mineral processing units derived from King et al. (2012) to construct four test cases. Table 1 displays the specific equations.

Both case 1 and 2 correspond to a simplified magnetic separation unit, in which all machine variables are represented by a constant, ECD is used as a proxy of particle volume, and two minerals are present: one paramagnetic and the other of negligible magnetic susceptibility. The sole difference between the two cases is their resulting entropy (Schach et al., 2019; Velázquez Martínez et al., 2019) due to the used coefficients. This was varied to test the effect of separation efficiency on the accuracy of LR model predictions.

Case 3 is a hydrocyclone with a 20 µm cut-size. A standard model for hydrocyclone recovery to concentrate is used to simulate the true probabilities of each particle of reporting to concentrate. Referring to the corresponding equation in table 1 shows that this curve is a monotonous transformation (via the function $f(x) = 1 - e^{-x}$) of a linear function of log-transformed ECD. Thus LR, having a similar functional structure as the hydrocyclone equation (using the logistic function $f(x) = 1/(1 + e^{-x})$ for the monotonous transformation), can reasonably approximate the theoretical behavior of this unit, but not perfectly, as will be seen later.

Case 4 corresponds to a flotation unit in which each of the minerals has a different flotation rate constant (k) and influence of size on the recoverability (φ, King et al., 2012, eq. 9.126). To calculate φ, 7.75 and 5.9 µm were used as the size of maximum recovery (ε) for minerals 1 and 2, respectively, while 145 and 90 µm were used as the respective maximum particle sizes at which a particle can be floated without detachment (δ) for minerals 1 and 2. The residence time is kept constant at 5 minutes.

In order to assess the performance of the method, the particle dataset was divided into train and test subsets following a 7:3 ratio. For each constructed case, a label was randomly assigned to each particle in the train subset according to its calculated

probabilities. A LR model was then fitted using the assigned particle labels. The predicted probabilities from this model can be compared directly with the “true” probabilities in the test subset (Fig. 2). Both train and test subsets are publicly available (Pereira et al., 2020).

As Figure 2 shows, the LR models were able to reconstruct the given probabilities in every case, albeit with better precision in cases 1, 2, and 3 than case 4 – the one with the most complex probability-defining equation. Even though the precision obtained in case four is lower, Figure 3 illustrates how the LR can reconstruct the approximately parabolic effect of particle size on recovery, given that ECD^2 is provided as an input variable (cf. section 2). Furthermore, figure 3 also illustrates how the LR captures the distinct flotation behavior of each mineral correctly. However, the fit is not perfect and this explains the deviations between calculated and predicted probabilities seen in Figure 2D.

The fitted model coefficients (Table 2) can also be examined to evaluate the capacity of the lasso regularization to omit non-relevant variables and assign meaningful coefficients to the ones incorporated into the model. In the simplest cases (1, 2, and 3), the LR-coefficients for the input variables roughly follow the magnitude and sign given in the defining equations. In addition, it every variable that was not given in the case-defining equation was omitted (Table 1).

However, the interpretation of coefficients in case 4 is more complicated since the assigned values depend on the main mineral (by mass) composing the particle. In this case, the magnitude of the coefficients related to mineral 1 are higher than the ones related to mineral 2 – an agreement with the case-defining equation. However, the regularization did not omit all variables that were not used in the equation (*aspect ratio and solidity*).

The failure of the models to consistently omit variables that were not used in the construction of case 4 may be attributed to the following factors:

1) **Covariance structure of the explanatory variables**. When some of the explanatory variances covary, it will generally be difficult to isolate the ones which are truly meaningful. For instance, particle shape may correlate strongly with particle size, such that it is also included in the model, even though it is not expressed in the defining equation.
2) **Non-linear physical relationships.** The functional form of LR models is restricted to linear combinations of the input variables to find the vector which best separates the different output classes within the space defined by the explanatory variables (Cox, 1958). In cases where the physical nature of the separation process is not described by a function with a similar structure, other variables may be included in the model to compensate for the effects of this non-linearity. This is seen clearly in case 4.

Overall, the constructed cases show that while LR models are clearly suitable for the reconstruction of probabilities, they are not a perfect tool for variable selection and fitted coefficients should therefore be analyzed with care.

3.2 Real case study

The practical applicability of the developed method is further demonstrated in a real case study on samples from the Boa Vista pyrochlore mine in Catalão, Brazil. The mine is part of the Catalão II carbonatite complex, one of the many intrusions of the Alto Paranaíba Igneous Province (Brod et al., 2000). The Boa Vista mine, owned and operated by CMOC International Brazil, commenced operations mining the weathered parts of the deposit but today exploits the fresh carbonatite. Pyrochlore is the main ore mineral, while phlogopite, calcite, dolomite, ankerite, apatite, feldspar, and magnetite constitute the main gangue minerals (Machado Junior, 1991; Palmieri, 2011).

The pyrochlore ore is processed in the Boa Vista Fresh Rock plant (BVFR). After comminution, the ore is treated with reverse carbonate flotation, reverse silicate flotation, magnetic separation, and finally pyrochlore flotation (Fig. 4). To demonstrate the adaptability of the proposed method, it is applied to reconstruct the continuous operation of the complete processing plant.

A sampling campaign was executed on the processing plant covering the feed of the first beneficiation unit (carbonate flotation), the products of every intermediate processing step, and the final tailings (Figure 4). The comminution and classification circuit was disregarded. Even though each flotation set-up has several units organized in rougher, cleaner and scavenger banks, the samples taken represent the total concentrate or tailings of the entire set-up. Thus, no attention is given to the internal streams of the flotation set-ups and each of the trained models represents the entire circuits encompassed by the darker boxes in Figure 4. This sampling strategy allowed us to use a minimal sample number without compromising too much on detail. The samples provided for the study were composites, taken under steady-state conditions, but without following the unit's residence time. The sample preparation and automated mineralogy analytical procedures are described in Pereira et al. (2019).

The concentrate and tailings streams of each circuit were used to train a LR model representing each processing unit. Summing up the particles of every concentrate and tailings samples, 1,162,090 particles described by 203 variables were used in the training phase. Afterwards, the plant feed (i.e. feed to carbonate flotation) was processed through the virtual processing plant following its real setup, and having the particles assigned as tailings in each step being fed to the next virtual unit operation until the final Nb concentrate is obtained (Figure 4). The feed sample contained 118,995 particles. The virtual process was bootstrapped 1000 times. Although the whole process was done without any mineral grouping, the set of 29 identified minerals was later grouped into 8 mineral groups to facilitate visualization/comparison of the process outcomes (Table 3). The distribution of every variable in every sample used to train the logistic regression models is aggregated, in terms of percentiles, in the Supplementary material 1.

For visual assessment, the predicted characterization results in each bootstrap run were divided by the actual results observed in the processing plant. The distribution of these ratios (log-transformed, base *e*) is displayed as box-plots in Figure 5. Clearly, the predicted results overlap with or are very close to 0 in most cases, indicating that the method provided accurate predictions of the mineral contents in every stream. The masses, however, are systematically incorrect in the carbonate and pyrochlore

flotation units. The mass is calculated from the prior adjustment step, therefore is mostly influenced by the balance of particle types (concentrate- and tailings-like) in the feed. Considering that the sampling scheme did not account for unit residence time, it might be that the proportion of particle types was affected by variations in feed composition due to the natural variability of the ore. This may have led to the incongruence between the simulation results and the real masses. Yet, since our method accurately assigned particles to the concentrate or tailings streams, the product compositions remained stable.

Another interesting observation is that the uncertainties for each variable increase towards the final processing units. This is because fewer particles reach the last processing units, and the random assignment of particles to output streams therefore has a greater impact on product composition than for the first steps of the process. A similar effect is seen for the minor components of individual streams. Mineral groups with large predicted uncertainties in specific streams (e.g. pyrochlore in the magnetic separation circuit or sulfides in the pyrochlore flotation – Figure 5) are generally minor constituents of these streams (Figure 6).

For a further physical comparison between the predicted and real results, Figure 6 displays the agreement between real and predicted mass flows of each mineral group in each stream. The median of the bootstrapped predictions is used to represent its obtained product, while the minimum and maximum error bars are represented by the 2.5 and 97.5 quantiles, respectively. Since only one sample of each stream was analyzed with the automated mineralogy system, no confidence interval could be plotted for the real masses.

Table 4 presents the entropy, number of omitted variables, and the most influential variables in each of the process units. Following the entropy calculation, the pyrochlore flotation and magnetic separation units are significantly more efficient than the carbonate and silicate flotation units are. In almost every case, the system omitted almost half of the total number of variables. For interpreting the coefficients, high and low values indicate the influence on reporting to the concentrate and tailings, respectively. Our method clearly selected some relevant variables in each case (e.g. apatite and strontianite surface content in the carbonate flotation, or the modal content of pyrochlore in the pyrochlore flotation). However, some coefficient assignments appear questionable, for instance in the magnetic separation unit where the modal content of siderite positively influences a particle's deportment to the concentrate, while the surface content of the same mineral has the opposite influence. The complete list of coefficients assigned to each variable in each separation unit is available in the Supplementary material 2.

Furthermore, the obtained particle probabilities are meaningful to identify optimization strategies for processing units. For this, the particles should be grouped according to their main mineral component. The distribution of probabilities of particles in each group can then be plotted in the form of a histogram. The efficiency of a process unit increases with the certainty that each mineral group will mostly report to one of its output streams. Within the proposed histogram, efficiency would be illustrated by a narrow distribution of probabilities of each mineral group, either lying close to 0 or to 1. The recovery of the carbonate minerals in the carbonate flotation is used here as an

example (Figure 7 A). Even though most particles of this mineral group have a high probability to report to the carbonate-rich product, a skewed distribution with many particles showing low recovery probabilities is noticeable. Figure 7 B illustrates that carbonate minerals have high recovery probabilities up to 35 µm. Siderite and magnesite_Fe are the only minerals from the group with low overall recovery probability. Studies on the effects of cell hydrodynamic conditions or regrinding would be key to optimize the recovery or carbonates to the carbonate-rich product.

4 Discussion and final considerations

As shown in the previous section, our method greatly enhances currently available approaches for particle tracking by successfully using complete high-dimensional particle datasets, with minimum human-input, to provide accurate estimations of the probabilities of each particle in a process unit to report to a specific output stream. Using the estimated probabilities, the method not only allows for accurate and detailed process prediction, but also opens up new possibilities for the understanding and optimization of minerals processing units. Below we discuss some aspects of the achieved results in more detail, and provide an outlook on potential future applications of particle tracking methods.

We start by noting that good results were achieved for the real case study despite the fact that we used a 2D dataset to describe 3D objects. This confirms our initial expectation that stereological effects, as long as they are the same across the entire dataset, will not strongly affect simulation outcomes (cf. section 2). Rather, they introduce additional statistical uncertainty. This may become a problem in cases where high precision is required for process prediction. In such cases, the acquisition of larger data volumes may provide the necessary detail. However, the extent to which this is possible may be limited by available computational capacity. Previous investigations (Miller et al., 2009; Ueda et al., 2018) suggest that the use of 3D-based particle data (e.g. provided by X-Ray computed tomography) could also improve results. We note in this context, that our method can readily incorporate such 3D particle datasets. It is not limited by the kinds of data that are currently available, and could even be used in future studies to investigate the effects of stereological bias in more detail.

Second, the lasso regularization successfully selected physically meaningful variable coefficients in both the constructed and real case studies. In addition, it omitted most of the other variables, providing relatively sparse predictive models. This indicates the efficiency of the method in reducing the dimensionality of a particle dataset without human supervision. However, it is also clear from the specific results that there is substantial complexity in the relationship of the fitted coefficients with actual variable importance, mostly due to using an analysis of covariance model. For instance, we already described above the difficulties in understanding the influence of specific variables for the constructed case 4. This case contained only two minerals. The difficulties in interpretation are expected to increase substantially in real ores composed of many more distinct minerals.

At least in part, these difficulties are due to the difference in measurement scales between the different variable classes (e.g. size and compositional variables). For instance, a naïve interpretation of raw coefficient values would suggest that particle mineralogy is more important than particle size for all units of the Catalão processing

plant, since the corresponding coefficient values are the largest (cf. Table 4). Fitted coefficients of *ECD* and ECD^2 are only intermediate in size (Table 4). However, the strong influence of particle size on the recovery of carbonate minerals in the carbonate flotation unit is clearly illustrated in Figure 7. Given this observation, we would suggest the use of plots such as those shown in Fig. 7, as the major tool to understand the influence of specific particle properties on process outcomes. Maps of average recovery probability as a function of two particle properties, similar to those presented in Schach et al. (2019; Fig. 11), could also be compiled.

Third, we note in this context that particle distribution-based models for minerals processing machinery were first developed from empirical observations of individual process units that were later generalized to quantitative physical models of mineral and particle behavior (King et al., 2012). The extensive application of our method in targeted case studies, followed by a complete exploration of the obtained particle probabilities, could be the next step to achieve the same development for particle datasets.

Fourth, we would like to point out that our method can be extended to provide an interface between particle and machine properties. This can be achieved by sampling a process under variable operating conditions, allowing for the quantification of the impact of these variations on the recoverability of individual particles, and the identification of optimum operating conditions given the process goals. In combination with the fact that our method provides reliable process predictions even when variable feed compositions are considered, this could be used to virtually assess plant performance prior to the processing of new feed, provided the material falls within the same parameter space as the original training data. Such a virtual assessment would in turn enable the fine-tuning of operational parameters to guarantee efficient operation, enabling resource and energy savings, and thus cleaner production.

Also in the framework of future applications, we should contemplate the relevance of this method in brownfield exploration. Considering an ore deposit in the vicinities of an operating processing plant, which can be assumed to belong to the same geological framework, this method could be used to evaluate the significance of its exploitation and processing in the existing facilities. However, this prospective application would require laboratory comminution tests that mimic the industrial operations since the method cannot yet forecast comminution processes. Moreover, dedicated test work should be done in parallel to validate the modelling results.

Altogether, the versatility of our method allows for its direct application in diverse cases and for varying purposes. Its predictive power makes it a suitable basis for improving plant control systems – an essential pre-condition for dealing with complex ores. Previous particle-based studies that lacked precise estimations of particle recoverability (e.g. Büttner et al., 2018; Pereira et al., 2019) can now be updated with this method for more accurate results. Using the entire information contained in particle datasets is the key to increasing the resource efficiency of the mining industry of the future, and thus advance towards a sustainable circular economy.

5 Acknowledgments

The authors would like to thank the German Federal Ministry for Education and Research (BMBF) for funding this project within the CLIENT-II initiative (Grant number 033R189B). We also thank our colleagues Edgar Schach, Marius Kern, Markus Buchmann, Juho Hannula, and Ivan Fernandes for fruitful discussions during method development, Thomas Heinig and Sabine Gilbricht for the help with the analytical work, and Peter Boelens for comments on the text.

6 References

Barry, S.C., Welsh, A.H., 2002. Generalized additive modelling and zero inflated count data. Ecol. Modell. 157, 179–188. https://doi.org/10.1016/S0304-3800(02)00194-1

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1007/9781441993267_5

Brod, J.A., Gibson, S. a., Thompson, R.N., Junqueira-Brod, T.C., Seer, H.J., Moraes, L.C. De, Boaventura, G.R., 2000. The Kamafugite-Carbonatite Association in the Alto Paranaíba Igneous Province (APIP) Southeastern Brazil. Rev. Bras. Geociências 30, 408–412. https://doi.org/10.25249/0375-7536.2000303408412

Büttner, P., Osbahr, I., Zimmermann, R., Leißner, T., Satge, L., Gutzmer, J., 2018. Recovery potential of flotation tailings assessed by spatial modelling of automated mineralogy data. Miner. Eng. 116, 143–151. https://doi.org/10.1016/j.mineng.2017.09.008

Cárdenas, E., 2017. Particle tracking in geometallurgical testing for Leveäniemi Iron ore , Sweden. Lulea Univ. Technol.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16. ACM Press, New York, New York, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Cox, D.R., 1958. The regression analysis of binary sequences. J. R. Stat. Soc. Ser. B 20, 215–242.

De Vasconcelos, M.J.P., Silva, S., Tome, M., Alvim, M., Pereira, J.M.C., 2001. Spatial prediction of fire ignition probabilities comparing logistic regression and neural networks. Photogramm. Eng. Remote Sensing 67, 73–81.

Fandrich, R., Gu, Y., Burrows, D., Moeller, K., 2007. Modern SEM-based mineral liberation analysis. Int. J. Miner. Process. 84, 310–320. https://doi.org/10.1016/j.minpro.2006.07.018

Frenzel, M., Bachmann, K., Carvalho, J.R.S., Relvas, J.M.R.S., Pacheco, N., Gutzmer, J., 2019. The geometallurgical assessment of by-products—geochemical proxies for the complex mineralogical deportment of indium at

Neves-Corvo, Portugal. Miner. Depos. 54, 959–982. https://doi.org/10.1007/s00126-018-0849-6

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Software, Artic. 33, 1–22. https://doi.org/10.18637/jss.v033.i01

Gaddum, J.H., 1945. Lognormal Distributions. Nature 156, 463–466. https://doi.org/10.1038/156463a0

Godinho, J.R. a., Kern, M., Renno, A.D., Gutzmer, J., 2019. Volume quantification in interphase voxels of ore minerals using 3D imaging. Miner. Eng. 144, 106016. https://doi.org/10.1016/j.mineng.2019.106016

Hannula, J., Kern, M., Luukkanen, S., Roine, a., van den Boogaart, K.G., Reuter, M. a., 2018. Property-based modelling and simulation of mechanical separation processes using dynamic binning and neural networks. Miner. Eng. 126, 52–63. https://doi.org/10.1016/j.mineng.2018.06.017

Hastie, T., Tibshirani, R., Wainwright, M., 2015. Statistical Learning with Sparsity, Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC. https://doi.org/10.1201/b18401

Heinig, T., Bachmann, K., Tolosana-Delgado, R., Boogaart, G. Van Den, Gutzmer, J., 2015. Monitoring gravitational and particle shape settling effects on MLA sampling preparation, in: IAMG Conference 2015. pp. 200–206.

Henderson, A.R., 2005. The bootstrap: A technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. Clin. Chim. Acta 359, 1–26. https://doi.org/10.1016/j.cccn.2005.04.002

Hodouin, D., 2011. Automatic Control in Mineral Processing Plants: an Overview., IFAC Proceedings Volumes. IFAC. https://doi.org/10.3182/20091014-3-cl-4011.00003

Ilisei, A.M., Khodadadzadeh, M., Ferro, A., Bruzzone, L., 2019. An Automatic Method for Subglacial Lake Detection in Ice Sheet Radar Sounder Data. IEEE Trans. Geosci. Remote Sens. 57, 3252–3270. https://doi.org/10.1109/TGRS.2018.2882911

Jowett, A., 1986. An appraisal of partition curves for coal-cleaning processes. Int. J. Miner. Process. 16, 75–95. https://doi.org/10.1016/0301-7516(86)90076-1

Keppel, G., Wickens, T.D., 2004. Design and Analysis: A Researcher's Handbook. Prentice Hall.

Ketcham, R. a., Carlson, W.D., 2001. Acquisition, optimization and interpretation of X-ray computed tomographic imagery: applications to the geosciences. Comput. Geosci. 27, 381–400. https://doi.org/10.1016/S0098-3004(00)00116-3

Khodadadzadeh, M., Jun Li, Plaza, A., Ghassemian, H., Bioucas-Dias, J.M., Xia Li, 2014. Spectral–Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization. IEEE Trans. Geosci. Remote Sens. 52, 6298–6314. https://doi.org/10.1109/TGRS.2013.2296031

King, R.P., Schneider, C.L., King, E.A., 2012. Modeling and Simulation of Mineral Processing Systems, Second. ed. Society for Mining, Metallurgy, and Exploration, Englewood, California.

Kirchherr, J., Reike, D., Hekkert, M., 2017. Conceptualizing the circular economy: An analysis of 114 definitions. Resour. Conserv. Recycl. 127, 221–232. https://doi.org/10.1016/j.resconrec.2017.09.005

Lamberg, P., Vianna, S., 2007. A technique for tracking multiphase mineral particles in flotation circuits. XXII Encontro Nac. Trat. Minérios e Metal. Extrativa - VII Meet. South. Hemisph. Miner. Technol. 195–202.

Lotter, N.O., 2011. Modern process mineralogy: an integrated multi-disciplined approach to flowsheeting. Miner. Eng. 24, 1229–1237. https://doi.org/10.1016/j.mineng.2011.03.004

Machado Junior, D. de L., 1991. Geologia e aspectos metalogenéticos do complexo alcalino-carbonatítico de Catalão II (GO). Universidade estadual de Campinas.

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: An applied review. Int. J. Remote Sens. 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

McCoy, J.T., Auret, L., 2019. Machine learning applications in minerals processing: A review. Miner. Eng. 132, 95–109. https://doi.org/10.1016/j.mineng.2018.12.004

Miller, J.D., Lin, C.-L., Hupka, L., Al-Wakeel, M.I., 2009. Liberation-limited grade/recovery curves from X-ray micro CT analysis of feed material for the evaluation of separation efficiency. Int. J. Miner. Process. 93, 48–53. https://doi.org/10.1016/j.minpro.2009.05.009

Olson, E., 2013. Particle Shape Factors and Their Use in Image Analysis-Part 1: Theory. J. GXP Compliance 15, 85–96.

Palmieri, M., 2011. Modelo Geológico e Avaliacão de Recursos Minerais do Depósito de Nióbio Morro do Padre, Complexo Alcalino-carbonatítico Catalão II, GO.

Pascoe, R.D., Power, M.R., Simpson, B., 2007. QEMSCAN analysis as a tool for improved understanding of gravity separator performance. Miner. Eng. 20, 487–495. https://doi.org/10.1016/j.mineng.2006.12.012

Pereira, L., Birtel, S., Möckel, R., Michaux, B., Silva, A.C., Gutzmer, J., 2019. Constraining the Economic Potential of By-Product Recovery by Using a Geometallurgical Approach: The Example of Rare Earth Element Recovery at Catalão I, Brazil. Econ. Geol. 1–14. https://doi.org/10.5382/econgeo.4637

Pereira, L., Frenzel, M., Khodadadzadeh, M., Tolosana-Delgado, R., Gutzmer, J., 2020. Particle dataset for constructing mineral processing case studies. https://doi.org/10.14278/rodare.336

Petruk, W., 2000. Applied mineralogy in the mining industry, 1st ed. Elsevier Science, Ottawa, Canada.

R Core Team, 2017. R: A Language and Environment for Statistical Computing.

Ripley, B.D., 1996. Pattern recognition and neural networks, 1st ed. Cambridge University Press, Cambridge.

Saerens, M., Latinne, P., Decaestecker, C., 2002. Adjusting the Outputs of a Classifier to New a Priori Probabilities: A Simple Procedure. Neural Comput. 14, 21–41. https://doi.org/10.1162/089976602753284446

Sandmann, D., 2015. Method Development in Automated Mineralogy. TU Bergakademie Freiberg.

Schach, E., Buchmann, M., Tolosana-Delgado, R., Leißner, T., Kern, M., Gerald van den Boogaart, K., Rudolph, M., Peuker, U. a., 2019. Multidimensional characterization of separation processes – Part 1: Introducing kernel methods and entropy in the context of mineral processing using SEM-based image analysis. Miner. Eng. 137, 78–86. https://doi.org/10.1016/j.mineng.2019.03.026

Tromp, K.F., 1950. Notion de probabilité dans le lavage. Quelques observations sur la courbe de fréquence des égarés. Rev. l'industrie minérale XXXI, 61–74.

Tromp, K.F., 1937. Neue Wege für die Beurteilung der Aufbereitung von Steinkohlen. Glückauf 6, 125–131.

Ueda, T., Oki, T., Koyanaka, S., 2018. Experimental analysis of mineral liberation and stereological bias based on X-ray computed tomography and artificial binary particles. Adv. Powder Technol. 29, 462–470. https://doi.org/10.1016/j.apt.2017.11.004

Van den Boogaart, K.G., Tolosana-Delgado, R., 2013. Analyzing Compositional Data with R. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36809-7_1

Velázquez Martínez, O., Van Den Boogaart, K.G., Lundström, M., Santasalo-Aarnio, A., Reuter, M., Serna-Guerrero, R., 2019. Statistical entropy analysis as tool for circular economy: Proof of concept by optimizing a lithium-ion battery waste sieving system. J. Clean. Prod. 212, 1568–1579. https://doi.org/10.1016/j.jclepro.2018.12.137

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wilks, D.S., 2009. Extending logistic regression to provide full-probability-distribution MOS forecasts. Meteorol. Appl. 16, 361–368. https://doi.org/10.1002/met.134

Wills, B.A., Finch, J., 2015. Wills' Mineral Processing Technology: An Introduction to the Practical Aspects of Ore Treatment and Mineral Recovery. Elsevier Science.

7 Tables and Figures

Table 1: The probability formula used in each of the constructed cases. ECD stands for equivalent circle diameter, mainmineral for the mineral of higher mass in the particle, Min_i for modal content of mineral i, $Min_i.s$ for surface content of mineral i, p for probability, φ for size-related parameter for flotation recovery, k for flotation rate constant, ε for size of maximum recovery, and δ for maximum particle size at which a particle can be floated without detachment

Case	Probability equation
1	$x = 0.20 + 0.10\log(ECD) + 0.35Min_1$
2	$x = 0.05 + 0.001\log(ECD) + 0.90Min_1$
3	$p = 1 - exp^{\left[-0.693\left(\frac{\log(ECD)}{\log(20)}\right)^{1.7}\right]}$
4	$\varphi = 2.33\sqrt{\frac{0.5\varepsilon^2}{(ECD)^2}}exp\left(-\frac{0.5\varepsilon^2}{(ECD)^2}\right)\left(1-\left(\frac{ECD}{\delta}\right)^{1.5}\right)$ $p = 1 - exp^{-(5k\varphi)}$ $\begin{cases}\varepsilon = 7.75; k = 0.7;\ \ \delta = 145, if\ mainmineral = mineral\ 1\\ \varepsilon = 5.9; k = 0.1;\ \ \delta = 90, if\ mainmineral = mineral\ 2\\ p = 0, if\ \varphi < 0\end{cases}$

Table 2: Coefficients obtained for each variable in the constructed cases. Different coefficients are assigned to the size and shape variables according to the particle's main mineral

Applies to	***Variable***	***Case 1***	***Case 2***	***Case 3***	***Case 4***
Both minerals	Mineral 1 modal	1.39	5.24	0.00	0.00
	Mineral 2 modal	0.00	0.00	0.00	0.00
	Mineral 1 surface	0.00	0.07	0.00	-0.20
	Mineral 2 surface	0.00	0.00	0.00	0.00
Mineral 1	Intercept	-1.01	-2.62	-2.63	-17.74
	Aspect ratio	0.00	0.00	0.00	-0.03
	Solidity	0.00	0.00	0.00	-1.87
	ECD	0.34	0.00	0.85	14.11
	ECD^2	0.00	0.00	0.00	-2.36
Mineral 2	Intercept	-1.01	-2.62	-2.63	-12.82
	Aspect ratio	0.00	0.00	0.00	-0.05
	Solidity	0.00	0.00	0.00	0.53
	ECD	0.34	0.00	0.85	9.44
	ECD^2	0.00	0.00	0.00	-1.78

*Table 3: Mineral groups with each of its members and their formulas. *The MLA analysis was not setup for distinguishing between hematite and magnetite, thus this mineral group might contain both species even though the mineral formula of hematite is indicated on the table*

Group	Mineral	Mineral formula
Carbonates	Calcite	$CaCO_3$
	Siderite	$Fe^{2+}CO_3$
	Strontianite	$SrCO_3$
	Ankerite	$Ca(Fe^{2+},Mg,Mn)(CO_3)_2$

	Magnesite_Fe	$(Mg,Fe)CO_3$
	Dolomite	$CaMg(CO_3)_2$
	Siderite_Mn	$(Fe^{2+},Mn)CO_3$
	Dolomite_FeSr	$CaMgSr(CO3)_2$
	Barytocalcite	$BaCa(CO_3)_2$
Rare earth minerals (REM)	Ancylite	$SrCe(CO_3)_2(OH)•(H_2O)$
	Calcioancylite	$CaCe(CO_3)_2(OH)•(H_2O)$
	Cebaite	$Ba_3Ce_2(CO_3)_5F_2$
Apatite	Apatite	$Ca_5(PO_4)_3(OH,F,Cl)$
Phlogopite	Phlogopite	$KMg_3Fe(Si_3Al)O_{10}(F,OH)_2$
Pyrochlore	Pyrochlore_Ca	$(Na,Ca)_2Nb_2O_6(OH,F)$
Iron and titanium oxides (FeTiO)	Iron oxide*	Fe_2O_3
	Ilmenite	$Fe^{2+}TiO_3$
	Rutile_Nb	TiO_2
Other silicates	Quartz	SiO_2
	Amphibole	$(Na,K,Ca)(Na,Mg,Fe^{2+},Mn^{2+},Ca)(Na,Mg,Fe^{2+},Mn^{2+},Ca)Si_8O_{22}$
	K-feldspar	$KAlFeSi_3O_8$
	Pyroxene_Ca	$Ca(Mg,Fe)(Si_2O_6)$
	Albite	$NaAlSi_3O_8$
	Sorosilicate	$Ca_2Al_2(Fe^{3+};Al)(SiO_4)(Si_2O_7)O(OH)$
	Pyroxene_Mg-Fe	$(Mg,Fe,Ca,Ti,Al)_2Si_2O_6$
	Hornblende	$(Ca,Na)_{2–3}(Mg,Fe,Al)_5(Al,Si)_8O_{22}(OH,F)_2$
Sulfides	Barite	$BaSO_4$
	Pyrite	FeS_2
	Chalcopyrite	$CuFeS_2$

Table 4: Calculated entropy of each processing unit, together with the number of variables that have not been used and the five variables of highest and lowest coefficients assigned by the LR model. High and low coefficients indicate influence to be deported to the concentrate and tailings, respectively. Mineral names followed by ".s" mean its content on a particle's surface

Product	Entropy	N° omitted variables	Highest coefficients	Lowest coefficients
Carbonate flotation	0.54	133	Apatite.s, Strontianite.s, Calcite.s, Dolomite_FeSr.s, Barytocalcite.s	Solidity.Apatite, Fe_oxide.s, Solidity, K.feldspar.s, Phlogopite.s
Silicate flotation	0.60	133	Phlogopite.s, Ca_Pyroxene.s, Amphibole.s, K.feldspar.s, Quartz.s	Solidity.Ca_Ancylite, Solidity.Quartz, Fe_oxide.s, Solidity.Phlogopite, Solidity.K.feldspar
Magnetic separation	0.14	140	Fe_oxide.s, Siderite, Mn.siderite, Pyrite, Ca_Pyroxene.s	Solidity.Fe_oxide, Solidity.Mn.siderite, Siderite.s, Intercept, Solidity
Pyrochlore flotation	0.07	122	Chalcopyrite.s, Ca_pyrochlore, Rutile_Nb.s, Pyrite, Solidity.Dolomite_FeSr	Solidity.Cebaite, Calcite.s, Strontianite.s, Solidity.Magnesite_Fe, Dolomite_FeSr.s

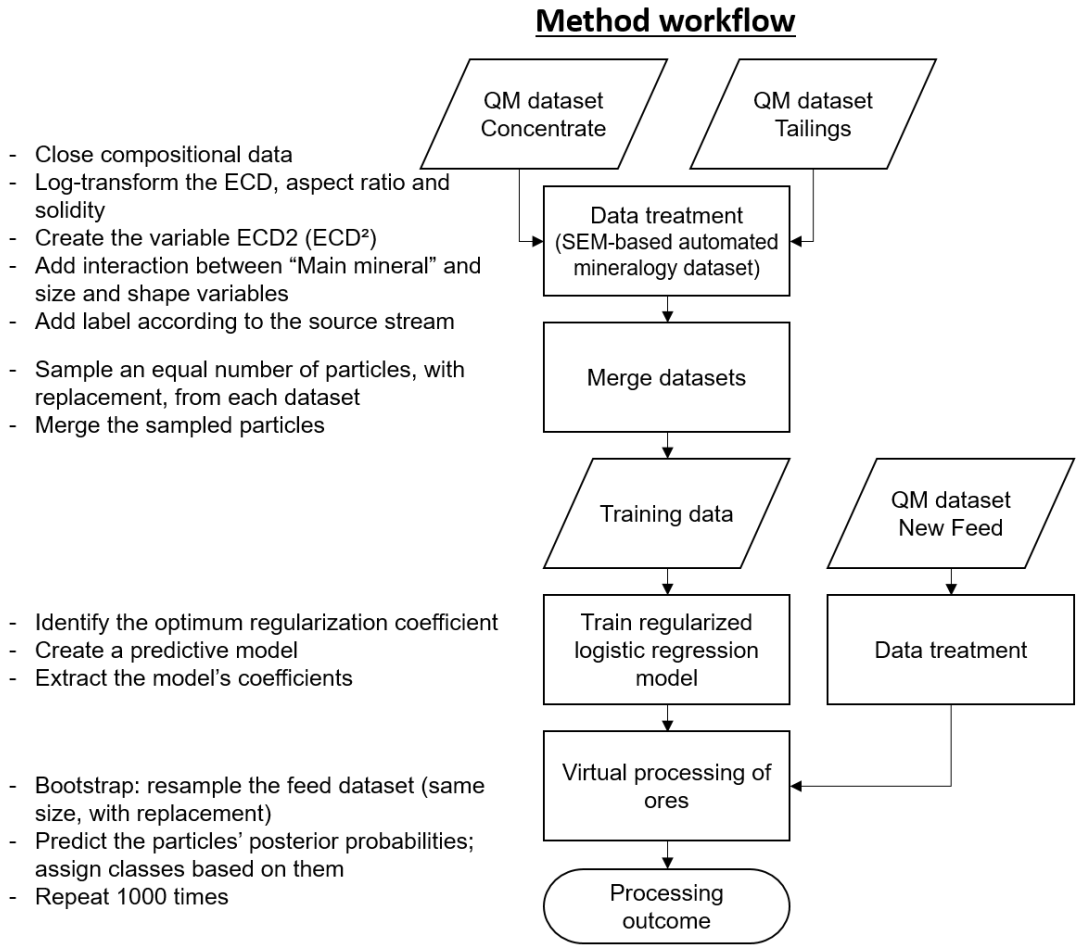

Figure 1: Workflow of method developed in this contribution. QM corresponds to quantitative mineralogy.

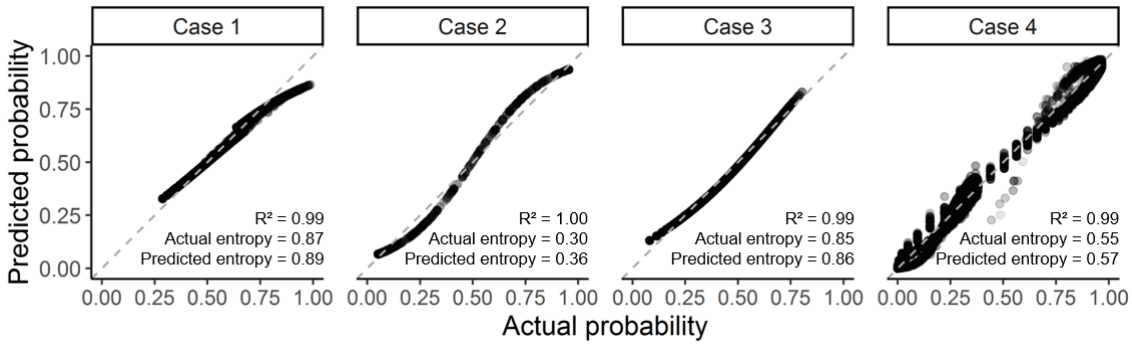

Figure 2: Fit of the probabilities obtained with the regularized logistic regression (LR) to the probabilities calculated for each case. In addition, the correlation between predicted and actual probabilities (R^2) and process entropy in each case.

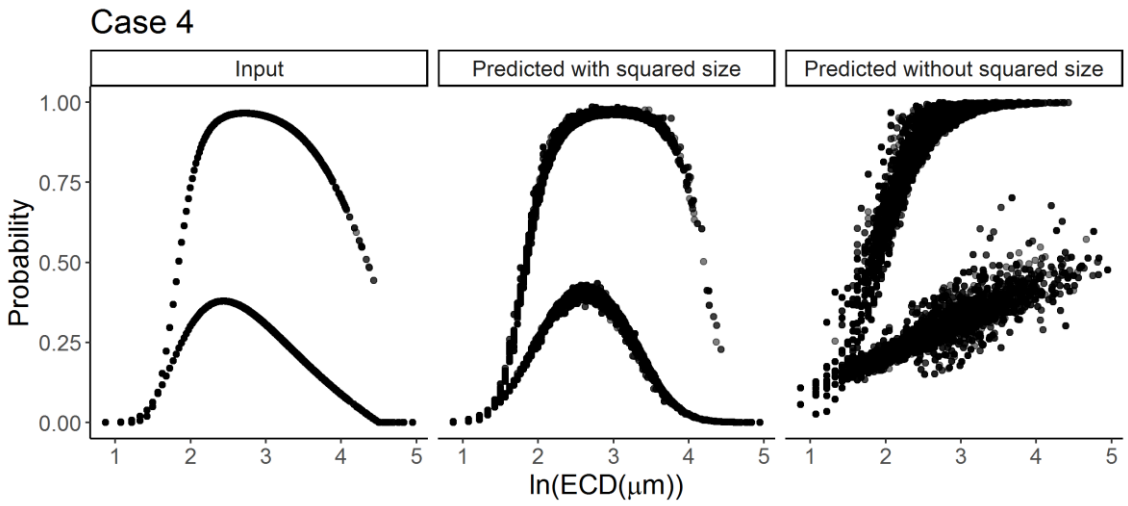

Figure 3: Aptitude of the lasso-regularized logistic regression on reconstructing the parabolic influence of a particle's size on its recoverability via flotation given that ECD^2 is provided as a variable (constructed case nº 4).

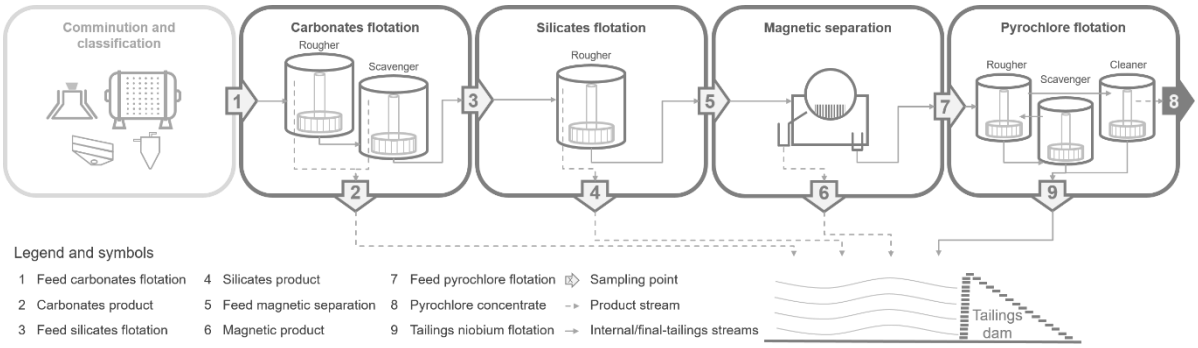

Figure 4: Simplified flowsheet of the Boa Vista Fresh Rock plant in Catalão, Brazil, with sampling points indicated. Each of the darker-outlined boxes corresponds to a model trained by the approach introduced here for virtually reconstructing the processing plant.

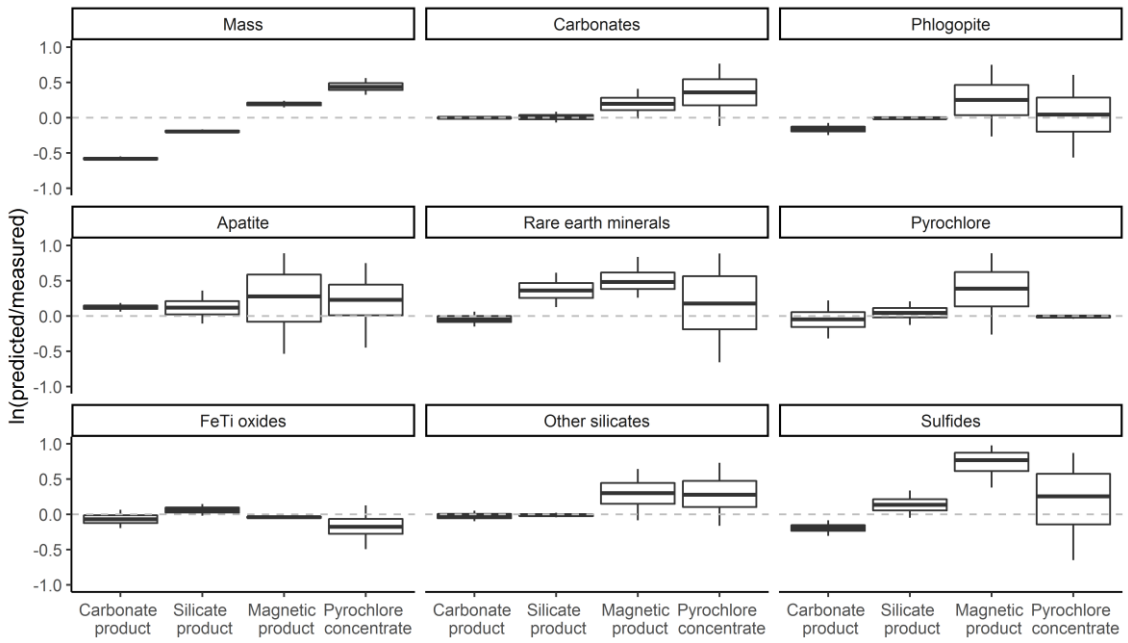

Figure 5: Distribution of the logarithm (base e) of the ratios between predicted and actual concentrate masses and compositions. The median, 0.25 & 0.75 percentiles, and 0.05 & 0.95 percentiles are represented by the center line, boxes, and whiskers, respectively.

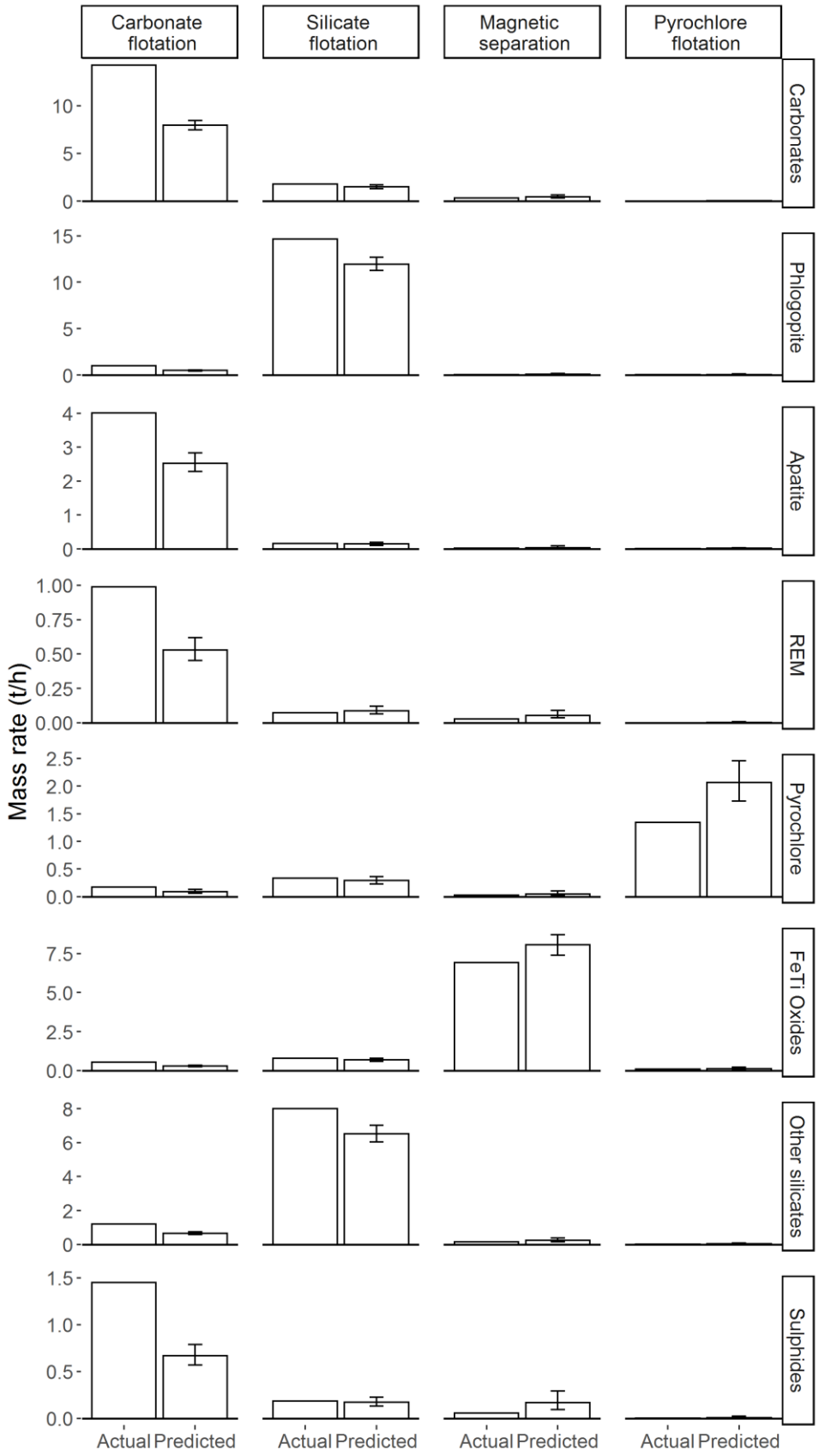

Figure 6: Median, minimum and maximum predicted and actual masses (t/h) of each mineral group in each concentrate stream.

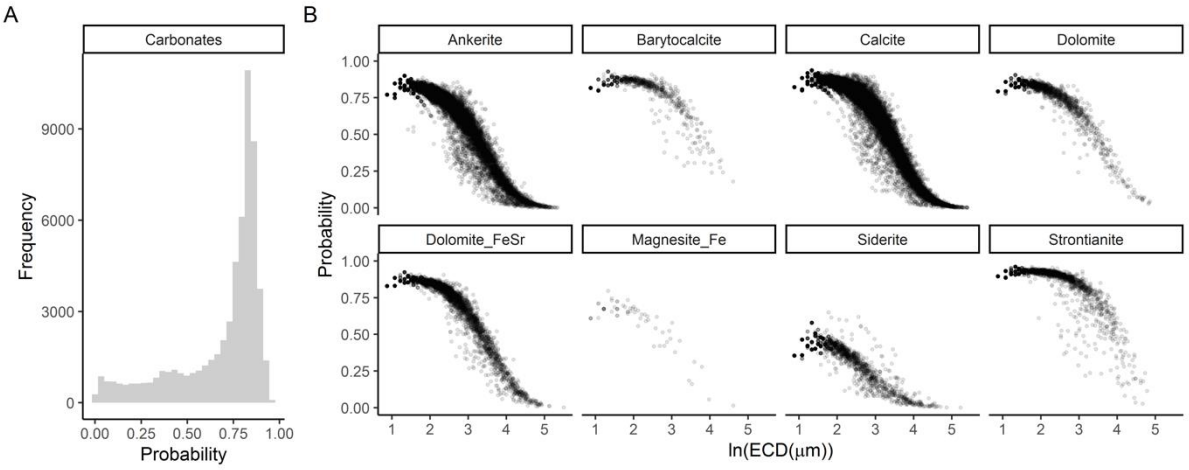

Figure 7: Probability histograms of particles in the carbonate flotation unit grouped by the main mineral comprising each particle. The distribution of particles among the mineral groups is indicated in the header boxes.