Ore Geology Reviews 82 (2017) 160–169

Contents lists available at ScienceDirect

Ore Geology Reviews

journal homepage: www.elsevier.com/locate/oregeo

A geostatistical approach to measure the consistency between geological logs and quantitative covariates

Amir Adeli [a,b,c], Xavier Emery [a,b,*]

[a] Department of Mining Engineering, University of Chile, Santiago, Chile
[b] Advanced Mining Technology Center, University of Chile, Santiago, Chile
[c] CSIRO-Chile International Center of Excellence in Mining and Mineral Processing, Santiago, Chile

ARTICLE INFO

Article history:
Received 21 October 2016
Received in revised form 21 November 2016
Accepted 22 November 2016
Available online 28 November 2016

Keywords:
Core logging
Geological domaining
Inconsistent data
Misclassification

ABSTRACT

Core logging is the geological study, recording and classification of petrophysical attributes of drill hole samples, such as lithology, alteration or mineralogical assemblage. The geological logging is qualitative and subject to errors because of its visual nature and other factors inherent to logging, such as low drill hole recoveries, difficulties in estimating the volumetric contents of minerals, or different logging criteria among geologists. To date, different tools for quality control and validation of geological logging have been elaborated, based on geological knowledge, statistics, geostatistics, image analysis, neural network and data mining. This paper presents an alternative approach based on geostatistical modeling for identifying and reclassifying potentially mislogged samples when quantitative covariates from geochemical analyses or metallurgical tests are available. The principle of this approach is to: (i) define geological domains for each quantitative variable by an adequate grouping of the log classes; (ii) transform the quantitative variables into normal scores, accounting for the previously defined domains, (iii) model the spatial correlation structure of the normal scores, (iv) perform leave-one-out cross validation and obtain predictions of the normal variables and the associated variance-covariance matrices of prediction errors; (v) calculate a measure of consistency for each sample and each possible logged class under a multivariate normal assumption; and (vi) compare these measures of consistency with the actual logged classes to detect suspicious logs. The methodology is demonstrated in a case study from an iron ore deposit, with data of rock type logged by geologists and seven quantitative variables (grades of elements of interest, loss on ignition and granulometry).

1. Introduction

Drilling is the most expensive procedure in mineral exploration. Information from drill holes can be extracted by different methods, such as assaying, down-the-hole geophysical logging or geological core logging (Knödel et al., 2007; Marjoribanks, 2010). The latter is the geological study, visual recording and classification of petrophysical attributes of drill cores (essentially extracted by diamond drilling), such as lithology, alteration or mineralogical assemblage. Relevant information gathered from geological core logging is the basis for constructing geological and geo-metallurgical models for mineral resources evaluation and classification, ore reserves definition and mine planning (e.g., Soltani and Hezarkhani, 2011). It is also often used for partitioning heterogeneous deposits into geological or geo-metallurgical domains in which the regionalized properties of interest can be interpreted as stationary fields (Sinclair and Blackwell, 2002; Moon et al., 2006; Yunsel and Ersoy, 2011; Haldar, 2013; Rossi and Deutsch, 2014).

However, due to the visual nature of logging, the classification of petrophysical attributes is qualitative and subject to errors, which may be explained by several factors (Manchuk and Deutsch, 2012; Cáceres and Emery, 2013): presence of complex rock textures caused by overprinting processes; inherent difficulty to estimate mineral percentages and thresholds; lack of geochemical analyses during logging; lack of experience of mining geologists; non-unique logging criteria among geologists; and low core recovery because mineralized and altered rock zones are frequently more fragile and are the first parts that are lost during coring. Inaccurate logs generate data that are inconsistent with geochemical analyses and metallurgical tests, such as low iron grades in supposedly supergene hematite zones or low acid consumption in supposedly calcareous rocks. Due to limited time

* Corresponding author at: Department of Mining Engineering, University of Chile, Santiago, Chile.
E-mail addresses: amir.adeli@amtc.uchile.cl (A. Adeli), xemery@ing.uchile.cl (X. Emery).

http://dx.doi.org/10.1016/j.oregeorev.2016.11.028
0169-1368/

and resources for relogging, inconsistent logs are generally seen as outliers or ignored in the geological or geo-metallurgical modeling stage (Theys, 1999; Cáceres and Emery, 2013).

Quality control and validation of geological logging have been discussed in the fields of geosciences and resources engineering, where manual to automated procedures have been elaborated, based on geological knowledge, statistics, geostatistics, image analysis, neural network and data mining tools (Hoyle, 1986; Agterberg, 1990; Luthi and Bryant, 1997; Taylor, 2000; Luthi, 2001; Hassibi et al., 2003; Bourgine et al., 2008, 2015; Ewusi and Kuma, 2011). Geological knowledge-based approaches are mainly focused on chronostratigraphy and sequencing methods for correlating the logs of different drill holes (Agterberg, 1990). Furthermore, geophysical well logging allows mapping of petrophysical signature and can be used for cross-checking geological logs (Spies, 1996; Hearst et al., 2000; Luthi, 2001; Dunn et al., 2002; Knödel et al., 2007; Soleimani et al., 2016). In recent years, Manchuk and Deutsch (2012) introduced a geostatistical measure of coherency for geological logs, based on a spatial correlation model, which can be applied for quality control of drill hole data. Another proposal by Cáceres and Emery (2013) relies on leave-one-out cross-validation in order to identify log data that disagree with quantitative measurements. This proposal was applied to a synthetic deposit characterized by a single quantitative variable (an assayed grade) and three logged classes (rock types) to which different logging errors were added. The results showed good ability to identify mislogged rock types based on grade information. However, the proposed approach is applicable to mineral deposits with hard boundaries between logged classes, i.e., when there is no spatial correlation among quantitative variables across the boundary between two classes, which is a too restrictive requirement for disseminated deposits.

This work aims at presenting an improvement of the approach by Cáceres and Emery (2013) for validating geological logs and identifying data for which the qualitative logged class is not in agreement with quantitative variables available from assays or metallurgical tests. The proposed methodology is also based on leave-one-out cross-validation, but will be set in a fully multivariate framework and will be applicable to mineral deposits where quantitative variables are correlated across boundaries between logged classes (soft boundaries). It relies on the following four working assumptions: (1) geochemical analyses, metallurgical tests and geological features such as lithology, alteration or mineralogical assemblage should be consistent in spatial and statistical terms, insofar as they are related responses to the same geological processes; (2) geochemical analyses and metallurgical tests are more accurate than geological logging; (3) a small proportion of the drill cores is likely to be mislogged; (4) the quantitative variables are regionalized and can be interpreted as realizations of spatial random fields that, within one or more domains that partition the deposit, are transforms of stationary Gaussian random fields, i.e., random fields whose finite-dimensional distributions are multivariate normal and are invariant under a translation in space. The latter assumption is the basis of most of the current geostatistical approaches used for simulating quantitative variables and for quantifying geological uncertainty in mineral deposits (Verly, 1983; Chilès and Delfiner, 2012; Rossi and Deutsch, 2014; Deutsch et al., 2016).

The paper is outlined as follows. Section 2 describes our proposal from a methodological point of view and leads to the definition of a statistical measure (p-value) for each drill hole sample that indicates how closely the classes of the logged petrophysical attribute agree with the quantitative variables resulting from geochemical analyses and metallurgical tests. Section 3 presents an application to drill hole samples of an iron ore deposit, for which information of the logged rock type and seven quantitative covariates is available. Conclusions follow in Section 4.

2. Methodology

2.1. Problem statement

Let us consider a set of drill hole samples with spatial coordinates $\{\mathbf{x}_1,\ldots,\mathbf{x}_A\}$, for which geological logs are available. These logs provide a class number (between 1 and K) for a petrophysical attribute (e.g., rock type). In addition, there are N quantitative variables (for example, grade assays) for the same samples as that of the geological logs. The information from the quantitative variables is supposed to be accurate, due to a proper quality assurance and quality control program, while that of the logs is subject to errors. The proposed methodology is explained in the following subsections and summarized in the schematic diagram presented in Fig. 1.

2.2. Geological domaining

For mineral resources evaluation in heterogeneous deposits, each quantitative variable is generally associated with geological domains that consist of a single log class or a group of log classes such that (i) the quantitative variable can be modeled as a stationary field (i.e., a field whose distribution is invariant under a spatial translation) within each domain and (ii) there is a change of distribution between one domain and another (e.g., one observes some discontinuity or some change in the spatial continuity when crossing the boundary between two domains). A contact analysis is helpful at this stage to determine the best grouping and therefore identify the geological domains (Glacken and Snowden, 2001; Rossi and Deutsch, 2014; Maleki and Emery, 2015). Also note that the domains can differ from one quantitative variable to another, depending on the geological characteristics that control the behavior of each variable; in other words, the geological controls may not be the same for all the quantitative variables.

2.3. Normal score transformation

The data of each quantitative variable within each of its geological domains are declustered in order to obtain a distribution corrected for possible biases caused by irregular sampling patterns, then normal score transformed, i.e., transformed into data that follow a standard Gaussian distribution (Deutsch and Journel, 1998; Chilès and Delfiner, 2012). The one-to-one relationships between the original variables and their normal scores can be stored in transformation tables.

As the geological domains associated with a given quantitative variable do not overlap, the normal score data associated with the different domains are totally heterotopic, i.e., they are not defined at the same locations. In contrast, the normal scores data associated with different quantitative variables may be isotopic, partially heterotopic or totally heterotopic, depending on whether the geological domains are the same, partially overlap or are disjoint and on the sampling design of the quantitative variables (isotopic or heterotopic).

2.4. Covariance analysis

Since sample cross-variograms cannot be calculated for heterotopic datasets (Wackernagel, 2003), the spatial correlation structure of the normal score data is inferred by calculating their sample direct and cross covariances. Direct covariances measure the spatial continuity of each quantitative variable within each of

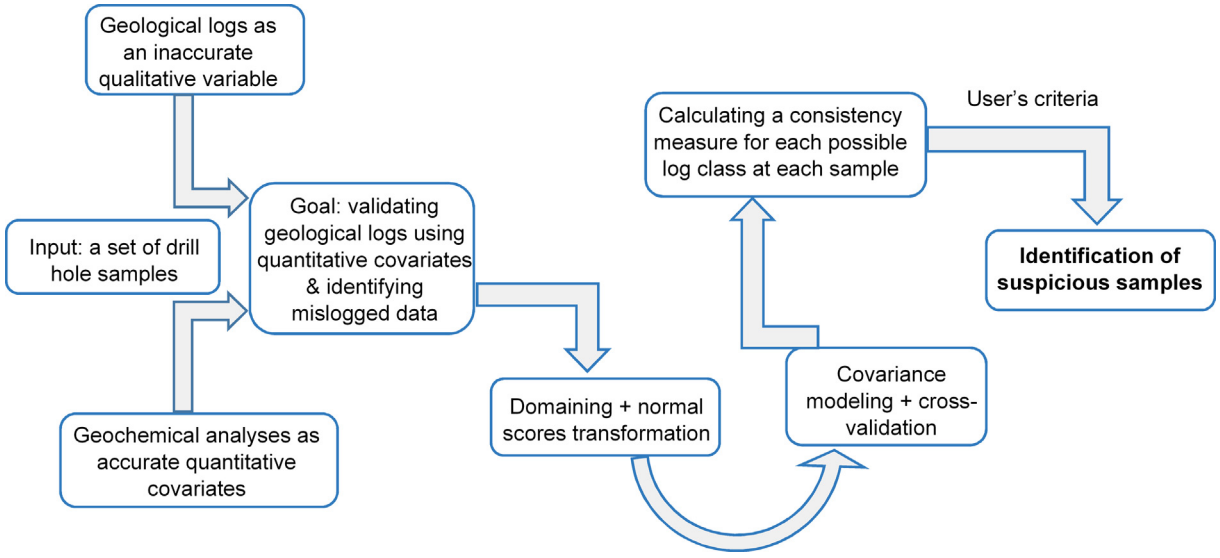

Fig. 1. Schematic diagram of proposed methodology.

its geological domains, while cross-covariances measure the spatial cross-correlation that exists between two different quantitative variables or the spatial cross-correlation of a quantitative variable between two different geological domains. This way, our methodology is able to account for soft boundaries, i.e., when measurements of a quantitative variable exhibit some correlation across domain boundaries (a frequent situation for disseminated deposits).

On the basis of the calculated sample covariances, a linear coregionalization model can be fitted by using combinations of basic nested covariance models (Wackernagel, 2003; Chilès and Delfiner, 2012), so as to define a set of theoretical direct and cross covariances for the Gaussian random fields that represent the quantitative variables within their different geological domains.

2.5. Cross-validation

Leave-one-out cross-validation (Chilès and Delfiner, 2012) is then performed in order to get a consistency measure between the data associated with the quantitative variables and the logged classes. Specifically, one drill hole sample is removed at a time from the dataset and all the Gaussian random fields at this sample are predicted by cokriging, by using all the remaining data located in the same geological domain and in the other domains. The results of cokriging for the sample under consideration are a vector of the prediction of the Gaussian random fields and a variance-covariance matrix of the prediction errors. This process is performed for each sample successively.

2.6. Calculation of p-values

For each sample $a \in \{1 \dots A\}$ and each class index $k \in \{1 \dots K\}$:

(1) Assume that, if no logging error occurred, class k would prevail at the sample under consideration.
(2) Transform the N quantitative variables measured at this sample into normal scores, using the suitable normal score transformation tables obtained in Subsection 2.3. Obtain a Gaussian vector $\mathbf{z}_{k,a}$ with N components.
(3) From the vector of cokriging prediction obtained in Subsection 2.5, extract the sub-vector $\mathbf{z}^*_{k,a}$ corresponding to the Gaussian random fields that are defined for the class under consideration (i.e., remove the prediction of the Gaussian random fields associated with geological domains that do not contain class k).
(4) Similarly, from the variance-covariance matrix of cokriging errors, extract the sub-matrix $\sum^*_{k,a}$ corresponding to the Gaussian random fields defined for the class under consideration.
(5) Based on the assumption made at stage (1), the conditional distribution of the Gaussian random fields at the sample location is multivariate normal with mean equal to the cokriging prediction and with variance-covariance matrix equal to that of the cokriging errors (Chilès and Delfiner, 2012). Accordingly, it has the following probability density function:

$$g(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma^*_{k,a})}} \times \exp\left\{-\frac{1}{2}(\mathbf{z}-\mathbf{z}^*_{k,a})^T (\Sigma^*_{k,a})^{-1}(\mathbf{z}-\mathbf{z}^*_{k,a})\right\} \quad (1)$$

where N is the common size of vectors $\mathbf{z}_{k,a}$ and $\mathbf{z}^*_{k,a}$ (number of original quantitative variables), $\mathbf{z}$ is a generic point of R^N, while $\mathbf{z}^*_{k,a}$ and $\sum^*_{k,a}$ have been obtained in the previous steps (3) and (4). Let us define the Mahalanobis distance as

$$r(\mathbf{z}) = \sqrt{(\mathbf{z}-\mathbf{z}^*_{k,a})^T (\Sigma^*_{k,a})^{-1}(\mathbf{z}-\mathbf{z}^*_{k,a})} \quad (2)$$

It provides a measure of the distance between the generic point $\mathbf{z}$ of R^N and the prediction $\mathbf{z}^*_{k,a}$ (expected value of the conditional distribution in Eq. (1)). It is dimensionless, scale-invariant and takes into account the correlations between the N Gaussian random fields defined for the class k under consideration. If $N = 1$, the distribution in Eq. (1) reduces to a univariate normal distribution and the Mahalanobis distance reduces to the standard score. For $N > 1$, the Mahalanobis distance is just a multidimensional generalization of the idea of measuring how many standard deviations away $\mathbf{z}$ is from $\mathbf{z}^*_{k,a}$.

The p-value $p_{k,a}$ associated with $\mathbf{z}_{k,a}$ (supposedly true Gaussian vector for sample a, which has been obtained at step (2)) is defined as the integral of the probability density in Eq. (1) over all the points of R^N with a Mahalanobis distance greater than $r(\mathbf{z}_{k,a})$. This

p-value gives a measure of the remoteness between the expected value $\mathbf{z}^*_{k,a}$ and the supposedly true value $\mathbf{z}_{k,a}$, i.e.:

$$\begin{aligned} p_{k,a} &= \int_{\mathbf{z} \text{ such that } r(\mathbf{z})>r(\mathbf{z}_{k,a})} g(\mathbf{z})\, d\mathbf{z} \\ &= 1 - \frac{1}{(2\pi)^{N/2}} \int_0^{r(\mathbf{z}_{k,a})} \exp\left\{-\frac{1}{2}t^2\right\} \frac{2\pi^{N/2}}{\Gamma\left(\frac{N}{2}\right)} t^{N-1}\, dt \end{aligned} \tag{3}$$

where $2\pi^{N/2}/\Gamma(\frac{N}{2})$ is the surface area of the unit-radius sphere of R^N (Chilès and Delfiner, 2012). To calculate $p_{k,a}$, let us define

$$I_N(r) = \int_0^r \exp\left\{-\frac{1}{2}t^2\right\} t^{N-1}\, dt \tag{4}$$

This integral can be computed by using integration by parts. One finds:

$$\begin{cases} I_1(r) = \sqrt{2\pi}\left(G(r) - \frac{1}{2}\right) \\ I_2(r) = 1 - \exp\left\{-\frac{r^2}{2}\right\} \\ I_N(r) = -r^{N-2} \exp\left\{-\frac{r^2}{2}\right\} + (N-2) I_{N-2}(r) \quad \text{for } N \geqslant 3 \end{cases} \tag{5}$$

where G stands for the standard Gaussian cumulative distribution function.

At the end of this process, one obtains, for each drill hole sample (a) and each class (k), a p-value $p_{k,a}$ that measures how "extreme" is the vector of N quantitative variables measured at this sample for the class under consideration. The analysis of the set of p-values will allow identification of the data for which the logged class is likely to be mistaken, as illustrated in the case study presented in the next section.

3. Case study

3.1. Presentation of the data

The proposed methodology is now applied to a data set from an iron ore deposit, the name and location of which are not disclosed for confidentiality reasons. The deposit is hosted by banded iron formations and is explored by diamond drill holes dipping from 60° to 90°. A total of 4096 samples are taken from cores and sent for analysis, yielding data for seven quantitative variables: grades of iron (Fe), silica (SiO_2), phosphorus (P), alumina (Al_2O_3) and manganese (Mn), loss on ignition (LOI) representing the mass concentration of volatile materials, and granulometric fraction of fragments with size above 6.3 mm (G). The dominant rock type is also available for each drill hole sample; it is deduced by geological logging and is classified into 10 classes:

- Code 1: Friable hematite (FH)
- Code 2: Compact hematite (CH)
- Code 3: Alumina-rich hematite (ALH)

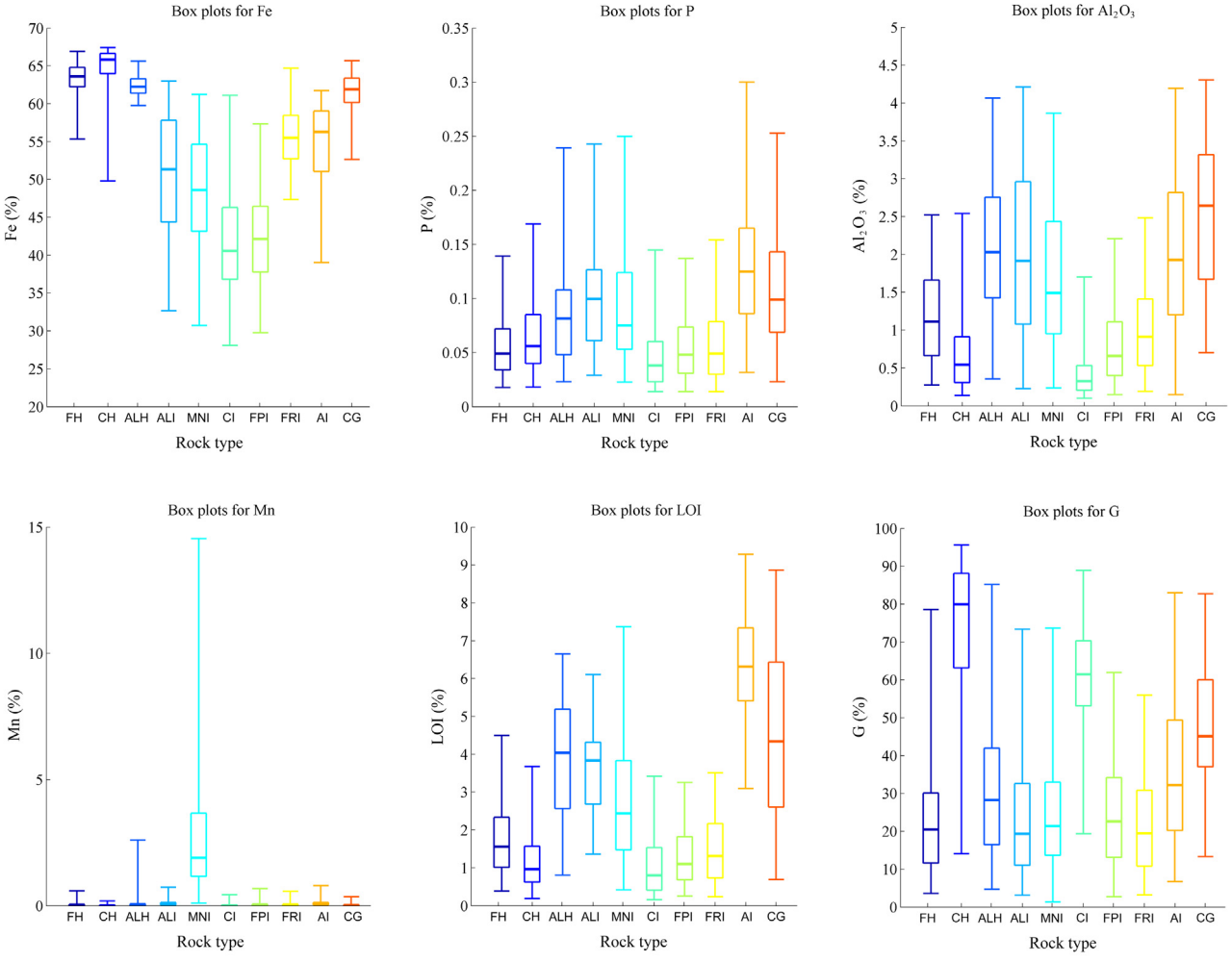

Fig. 2. Box plots showing quantiles Q2.5, Q25, Q50, Q75 and Q97.5 of the distributions of iron grade, phosphorus grade, alumina grade, manganese grade, loss on ignition and granulometry, for each rock type.

- Code 4: Alumina-rich itabirite (ALI)
- Code 5: Manganese-rich itabirite (MNI)
- Code 6: Compact itabirite (CI)
- Code 7: Friable iron-poor itabirite (FPI)
- Code 8: Friable iron-rich itabirite (FRI)
- Code 9: Amphibolitic itabirite (AI)
- Code 10: Canga (CG).

The deposit is divided into a supergene layer (surficial canga) and underlying ferruginous rocks. On the one hand, CG is a layer of well-consolidated rock composed mainly of goethite derived from weathering of the iron formation, with high iron grade, alumina grade and loss on ignition. On the other hand, the underlying rocks are subdivided on the basis of their granulometry and contents of iron, alumina, manganese and loss on ignition. Specifically, iron is mainly present in the form of hematite and itabirite. The former is an oxide-facies formation with high iron grade (most often above 62%), while the latter is a laminated, metamorphosed oxide-facies formation in which iron is present as thin layers of hematite, magnetite or martite, with grade generally less than 62% (Dorr, 1964). The alumina and phosphorus grades are high in ALH, ALI, MNI and AI, with the last two rock types also having high manganese grade and high loss on ignition, respectively. Finally, the compact rock types CH and CI exhibit a coarse granulometry, with values of G most often above 50%, while the other rock types are associated with finer granulometry (G mostly below 50%) (Fig. 2). These suggest that the quantitative variables are closely related to the rock type definition and that the proposed approach will provide an opportunity to detect inconsistencies between the former and the latter. In the following subsection, the methodology proposed in Section 2 is applied step-by-step. A general workflow is presented in Fig. 3.

Original data

Sample	Rock type	Fe	SiO_2	P	Al_2O_3	Mn	LOI	G
1	4	48.608	21.846	0.031	3.829	0.04	2.278	6.285
2	8	52.931	10.602	0.263	4.639	0.168	7.003	32.84
3	9	56.23	4.224	0.222	4.746	0.042	8.987	41.606
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
n = 4096	7	43.755	36.257	0.064	0.31	0.01	0.903	52.579

Domaining + normal scores transformation

Normal scores data

Sample	Rock type	Fe-1	Fe-2	Fe-3	SiO_2-1	.	.	.	LOI-3	G-1	G-2
1	4	-	-0.595	-	-	.	.	.	-	-	-1.377
2	8	-	-0.113	-	-	.	.	.	2.862	-	0.569
3	9	-	0.382	-	-	.	.	.	-	-	0.99
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
n = 4096	7	-	-	0.58	-	.	.	.	-0.321	-	1.429

Covariance modeling + cross-validation

Cross-validation results

		Estimates for										[19×19] Variance-covariance matrix
Sample	Rock type	Fe-1	Fe-2	Fe-3	SiO_2-1	.	.	.	LOI-3	G-1	G-2	Σ^*
1	4	0.283	0.013	0.365	-0.189	.	.	.	0.374	-1.177	-0.704	Σ_1^*
2	8	-0.015	0.686	1.249	0.136	.	.	.	0.379	-0.934	0.532	Σ_2^*
3	9	-0.022	0.566	0.800	-0.516	.	.	.	1.210	0.023	1.190	Σ_3^*
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
n = 4096	7	-1.007	1.395	0.936	1.396	.	.	.	-0.542	0.928	1.111	Σ_n^*

p-value calculation for assumed rock type

Assuming rock type = 1	= 2	= 3	= 4	= 5	= 6	= 7	= 8	= 9

Sample	Rock type	Estimates related to the assumed rock type = 1							[7×7] Variance-covariance matrix related to the assumed rock type = 1	***p*-value related to the assumed rock type = 1**
		Fe-1	SiO_2-1	P-3	Al_2O_3-3	Mn-2	LOI-3	G-2		
1	1	0.283	-0.189	-0.043	1.614	-0.012	0.374	-0.704	$\Sigma_{1,1}^*$	0.3095
2	1	-0.015	0.136	0.367	0.409	0.776	0.379	0.532	$\Sigma_{1,2}^*$	0.0000
3	1	-0.022	-0.516	0.882	0.869	0.286	1.210	1.190	$\Sigma_{1,3}^*$	0.0962
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
n = 4096	1	-1.007	1.396	0.254	-0.638	-0.727	-0.542	1.111	$\Sigma_{1,n}^*$	0.0011

User's criteria

Identification of suspicious samples

Fig. 3. Workflow of proposed methodology for identifying mislogged rock types.

3.2. Geological domaining

The first modeling step involves partitioning the deposit into domains in which the quantitative variables are homogeneously distributed and can be represented by stationary random fields (Rossi and Deutsch, 2014). To this end, for each quantitative variable, the rock types are grouped into domains based on the following considerations:

- The results of contact analysis, which aims at determining the behavior of the mean value of a quantitative variable in the neighborhood of the boundary between two rock type domains. Some examples are provided in Fig. 4, where the mean values are seen to change abruptly when crossing a rock type boundary, indicating that the rock types on either side of the boundary belong to different domains.
- The data statistics (mainly the mean value and the standard deviation) within each rock type. The rock types belonging to the same domain are expected to have similar statistics. An example is given in Fig. 5 for the iron grade and loss on ignition.
- The location and contact relationships of the rock types in the deposit. In particular, CG appears as a specific domain for all the quantitative variables because of its surficial position and because the data statistics in CG differ from those in any other rock type (Fig. 2).
- The lithological controls on the variable under study. For example, hematites and itabirites are likely to form two different domains for the iron and silica grades, but not necessarily for the other quantitative variables. Likewise, the compact and friable rock types are likely to define two different domains for the granulometry, but not for the remaining quantitative variables.

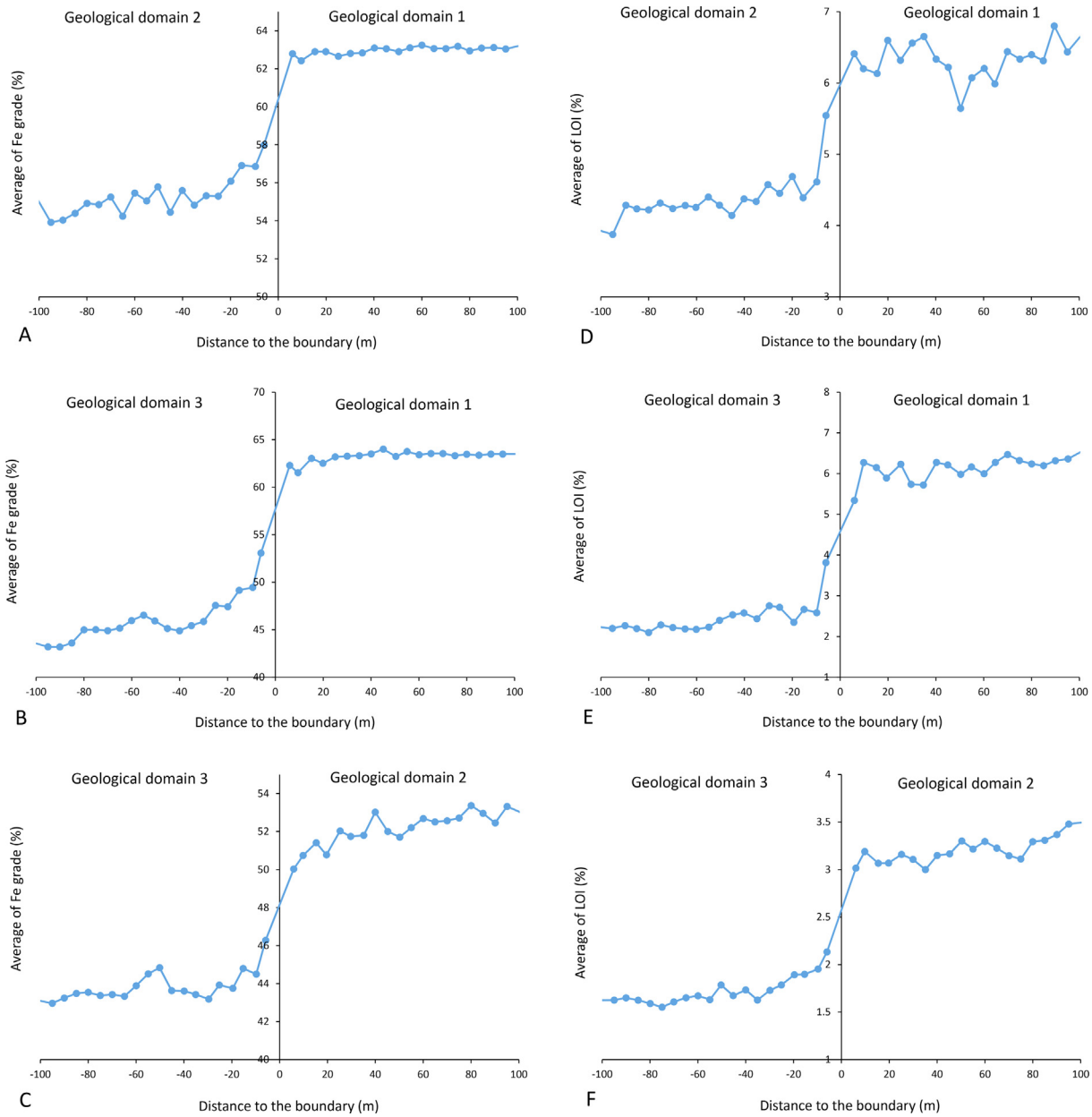

Fig. 4. Contact analysis for Fe (A, B and C) near the boundaries of geological domains 1 (FH + CH + ALH), 2 (ALI + MNI + FRI + AI) and 3 (CI + FPI), and for LOI (D, E and F) near the boundaries of geological domains 1 (AI), 2 (ALH + ALI + MNI) and 3 (FH + CH + CI + FPI + FRI).

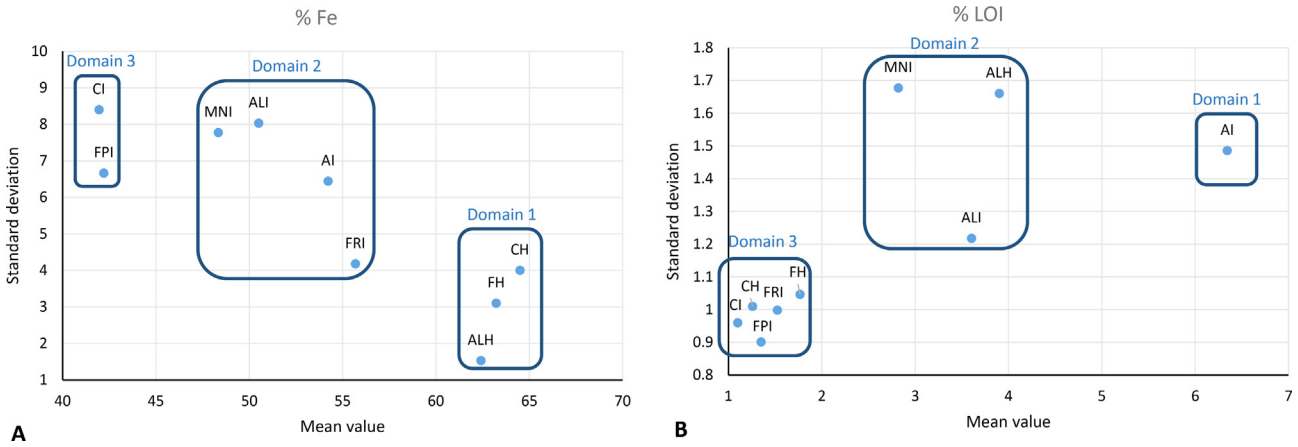

Fig. 5. Mean values and standard deviations of Fe (A) and LOI (B) for the underlying rock types (excluding canga), and grouping of these rock types into geological domains.

Table 1
Geological domaining for each quantitative variable.

	FH	CH	ALH	ALI	MNI	CI	FPI	FRI	AI
Fe	1	1	1	2	2	3	3	2	2
SiO_2	1	1	1	2	2	3	3	2	2
P	3	3	2	2	2	3	3	3	1
Al_2O_3	3	3	2	2	2	3	3	3	1
Mn	2	2	2	2	1	2	2	2	2
LOI	3	3	2	2	2	3	3	3	1
G	2	1	2	2	2	1	2	2	2

In the following, the study will focus on the underlying ferruginous rocks, i.e., we will no longer consider surficial canga, as its characterization depends more on the geographical position than on the quantitative variables. Table 1 gives a summary of the geological domains defined in the underlying rocks, where each quantitative variable is associated with two or three domains.

3.3. Normal scores transformation and coregionalization modeling

Overall, there are 19 combinations of quantitative variables and geological domains: Fe-1, Fe-2, Fe-3, SiO_2-1, SiO_2-2, SiO_2-3, P-1, P-2, P-3, Al_2O_3-1, Al_2O_3-2, Al_2O_3-3, Mn-1, Mn-2, LOI-1, LOI-2, LOI-3, G-1 and G-2. For each variable and each domain, the corresponding data are normal score transformed. The direct and cross covariances of the normal scores data are then calculated along the horizontal and vertical directions, identified as the main anisotropy directions, and fitted with a full linear coregionalization model thanks to a semi-automated fitting algorithm (Emery, 2010). The model uses a nugget effect and a set of nested exponential covariances with a geometric anisotropy between the horizontal and vertical directions, as its basic structures (Fig. 6).

Note that the normal scores associated with different quantitative variables and/or with different geological domains are assumed to be dependent, i.e., the linear coregionalization model assumes the existence of spatial cross correlations between all the 19 underlying Gaussian random fields. Accordingly, even if a quantitative variable is discontinuous across the boundary between two different domains, its values within these two domains are not necessarily independent (a situation sometimes referred to as a "transitional" boundary).

The above model parameters (normal score transformation tables and theoretical direct and cross covariances) will not change in all the subsequent stages of the study, insofar as we assumed that only a small proportion of the samples is mislogged, hence their effect on the model parameters should be marginal.

3.4. Cross validation

For each drill hole sample, the 19 Gaussian random fields are predicted by simple cokriging, using a vector of zero means, the previously fitted linear coregionalization model and a moving neighborhood (in the present case, a ball with a radius of 300 meters, divided into octants, in each of which the 4 closest samples are searched) as input parameters, and the normal scores available at these closest samples as input data. Cokriging yields a 19×1 vector of predictions and a 19×19 variance-covariance matrix of prediction errors for each drill hole sample.

3.5. Definition of consistency measures (p-values)

For each drill hole sample (index $a = 1,\ldots, 4096$) and each rock type (index $k = 1,\ldots, 9$):

- Assume that rock type k prevails at the sample under consideration.
- In view of the geological domains associated with the quantitative variables (Table 1), identify which of the 19 Gaussian random fields are defined for sample a.
- Transform the measured values of the original quantitative variables (grades, loss on ignition and granulometry) into normal scores, using the transformation tables associated with the adequate geological domains. Obtain the "true" Gaussian vector $\mathbf{z}_{k,a}$ of size 7×1. For example, for $k = 1$ (friable hematite FH), the Gaussian values associated with Fe-1, SiO_2-1, P-3, Al_2O_3-3, Mn-2, LOI-3 and G-2 are obtained, while the remaining ones are not.
- From the cross-validation results, extract the 7×1 vector of predictions ($\mathbf{z}^*_{k,a}$) and the 7×7 variance-covariance matrix of prediction errors ($\mathbf{\Sigma}^*_{k,a}$).
- Calculate the p-value $p_{k,a}$, as per Eq. (3).

3.6. Detection of suspicious data

To detect samples that may be mislogged, the following criteria based on the calculated p-values are considered:

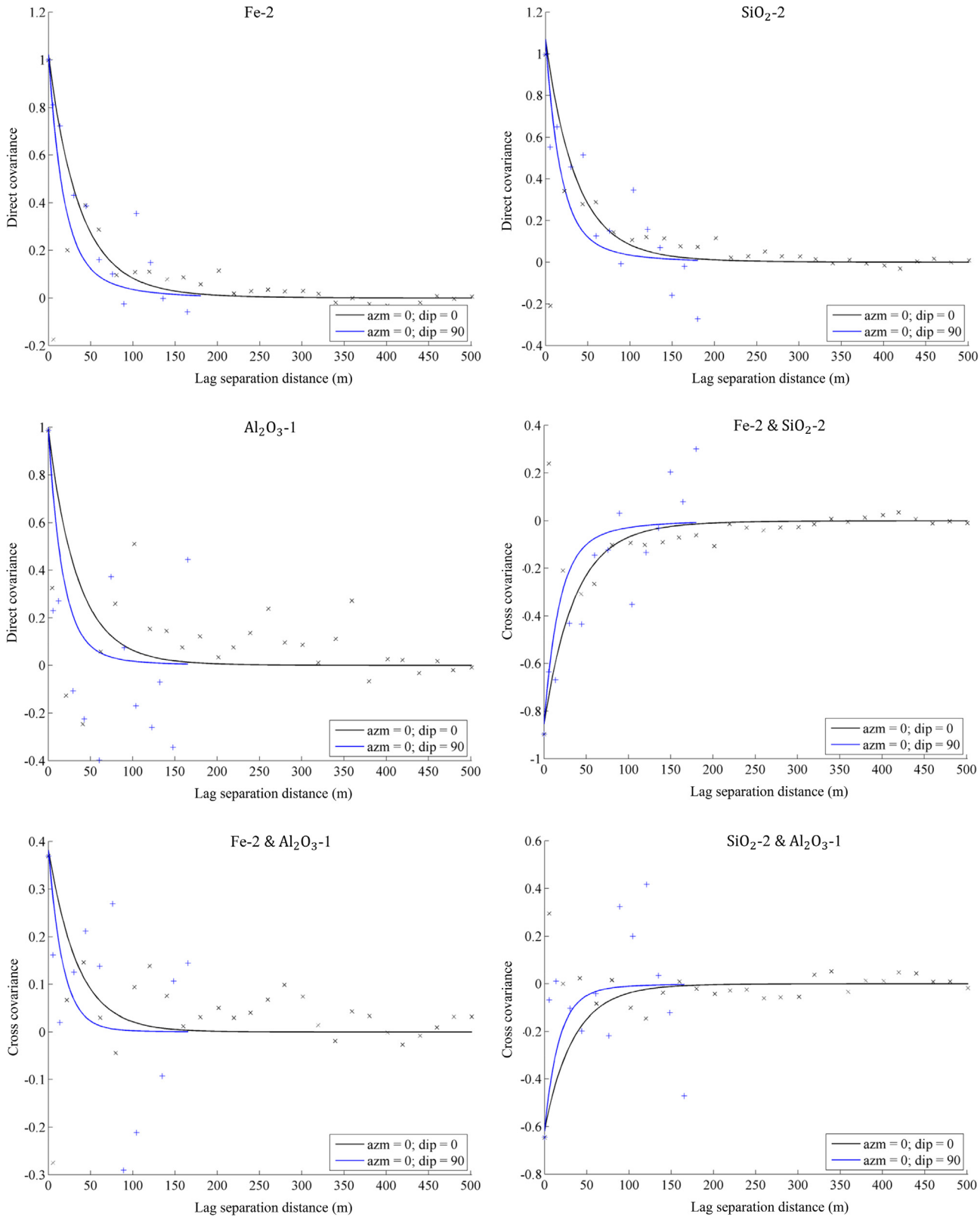

Fig. 6. An example of fitted direct covariances and cross-covariances for the transformed iron grade in domain 2 (Fe-2), transformed silica grade in domain 2 (SiO_2-2) and transformed alumina grade in domain 1 (Al_2O_3-1).

- The *p*-value associated with the class that has been actually logged by the geologist should be less than 0.01, i.e., the logged class is very unlikely.
- The *p*-value associated with some other class is more than 0.3 (quite likely).

In addition, to get results that are consistent with the geological zonation of the deposit, a geographical criterion is also considered, namely, that the alternative (likely) class has been logged at some drill hole sample distant less than 50 m from the suspicious sample.

Table 2
Geochemical assay results for specified suspicious samples.

Sample number	Fe	SiO_2	P	Al_2O_3	Mn	LOI	G
50	58.334	3.351	0.23	4.184	0.034	7.273	61.465
61	58.575	3.34	0.221	4.051	0.034	7.139	62.23
110	61.293	5.146	0.026	3.12	0.02	1.935	13.9
138	36.643	4.958	0.069	2.431	24.729	7.077	41.5
146	59.735	3.414	0.058	3.738	0.014	6.263	84.18
1910	62.103	2.395	0.151	1.011	2.878	2.453	13.691

Accordingly, each sample meeting all three mentioned conditions represents a suspicious data that should be examined on a case-by-case basis. Note that the geographical distance limit (50 m) and the *p*-value limits (0.01 and 0.3) can be freely modified by the practitioner, so that one could be more or less conservative in finding such suspicious data.

3.7. Results and discussion

The application of the previous criteria yields a total of 2.42% of the drill hole samples identified as suspicious. In the following, a few of these suspicious samples are discussed in more details, in the light of their logged rock types, values of the quantitative variables (grades, loss on ignition and granulometry) (Table 2), *p*-values (Table 3) and information of the 10 closest samples (Table 4).

Sample n°61: This sample has been logged by the geologist as rock type 3 (ALH), but the *p*-value of this rock type is close to 0 and the *p*-value of rock type 9 (AI) is very high (about 0.99) (Table 3). Looking at the assayed grades in Table 2, the measured iron grade appears to be abnormally low for an ALH, while the high loss on ignition agrees well with an AI. Therefore, it seems that sample n°61 is mislogged and should be relogged as rock type 9.

Sample n°50: This sample is the opposite case of sample n°61, as it has been logged as rock type 9 (AI), but the *p*-value of this rock type is close to zero and that of ALH (rock type 3) is close to 0.98 (Table 3). However, the measured grades (Table 2) are very similar to that of sample n°61 and agree well with rock type 9. Actually, sample n°61 (mislogged as rock type 3) turns out to be the closest one (Table 4), therefore the most influential in the cross-validation procedure, which explains why rock type 9 appears so likely. One concludes that sample n°50 is logged correctly and that the calculated consistency measures (*p*-values) are distorted due to the mislogged neighboring sample.

Sample n°110: This sample has been logged as rock type 8 (FRI), but our criteria suggest changing it to rock type 3 (ALH). Based on the assays (Table 2) and the existence of samples with rock type 3 in the neighborhood (Table 4), it seems that sample n°110 is effectively mislogged and should be changed to rock type 3.

Sample n°138: This sample has been logged as rock type 1 (FH), but may be relogged as 5 (MNI). This relogging would definitely agree with the measured manganese grade (more than 24%) (Table 2) and with the fact that the closest sample (sample n °137, less than 10 m away) is also logged as rock type 5 (Table 4). Note that another very close sample (n°1287) is logged as rock type 1 (Table 4), but the manganese grade measured for this sample is substantially lower (0.26%) and the iron grade substantially higher (60.95%), consistent with FH (code 1). Therefore, sample n°138 should be relogged as MNI (code 5), the same as its neighbor sample n°137, while sample n°1287 is correctly logged as rock type 1.

Sample n°146: This sample has been logged as rock type 1 (FH), but could be relogged as rock type 9 (AI). From Table 2, it seems that rock type 9 is more likely (owing to the high value of LOI) but not decisively (owing to the high value of Fe and low value of SiO_2, more compatible with a hematite). From Table 4, one observes that the closest sample (sample n°145) whose assays are very close to that of sample n°146 is coded as rock type 9,

Table 3
Calculated *p*-values for all possible rock types of specified suspicious samples.

Sample number	FH (1)	CH (2)	ALH (3)	ALI (4)	MNI (5)	CI (6)	FPI (7)	FRI (8)	AI (9)
50	0.316	0.292	0.983	0.865	0.000	0.006	0.007	0.114	0.003
61	0.027	0.021	0.000	0.000	0.000	0.001	0.001	0.069	0.988
110	0.006	0.001	0.881	0.143	0.010	0.000	0.001	0.000	0.062
138	0.000	0.000	0.000	0.000	0.304	0.000	0.000	0.000	0.000
146	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.913
1910	0.599	0.000	0.327	0.142	0.175	0.001	0.218	0.302	0.003

Table 4
Distance, sample number and logged rock type of the 10 closest samples to the specified suspicious samples.

Sample Number		1	2	3	4	5	6	7	8	9	10
50	Distance (m)	5.1	10.2	20.2	30.2	40.2	45.8	48.8	50.8	53.9	57.6
	Sample No.	61	936	1042	568	614	615	563	591	321	917
	Rock type	3	3	3	3	3	7	3	3	3	3
61	Distance (m)	5.1	5.1	15.1	25.1	35.1	40.7	44.8	49.9	52.3	53.5
	Sample No.	50	936	1042	568	614	615	563	591	321	917
	Rock type	9	3	3	3	3	7	3	3	3	3
110	Distance (m)	5.1	5.2	13.5	13.7	18.8	20.3	25.6	37.7	41.5	44.6
	Sample No.	2472	109	118	664	604	830	446	32	44	45
	Rock type	8	3	4	7	1	4	1	3	9	9
138	Distance (m)	6.1	6.1	14.3	16.1	25.9	43.0	50.6	58.8	191.3	191.4
	Sample No.	137	1287	721	661	1076	714	1026	1934	2169	1115
	Rock type	5	1	1	1	1	1	8	7	5	5
146	Distance (m)	5.3	7.2	11.8	17.2	18.2	27.2	34.0	40.9	49.2	50.1
	Sample No.	145	1373	516	2433	358	2370	3358	1328	13	979
	Rock type	9	1	3	1	3	2	2	3	3	3
1910	Distance (m)	6.9	9.8	19.7	21.8	29.0	31.8	35.7	39.6	43.9	46.9
	Sample No.	1942	2716	3126	2311	2043	3382	3257	3374	3182	1083
	Rock type	3	2	2	3	2	3	2	3	1	8

which may explain the *p*-value obtained for rock type 9. But, as for sample n°146, it cannot be said decisively whether rock type 9 or 5 is correct for sample n°145, so it is advised to physically check and relog these two samples.

Sample n°1910: This sample has been originally logged as rock type 2 (CH) but, according to the calculated *p*-values, it could be rock type 1 (FH), 3 (ALH) or 8 (FRI) (Table 3). From Table 2, it can be understood that codes 2 and 8 are not adequate for this sample (because of the too low granulometry and the too high iron grade, respectively). Finally, it is suggested to discard codes 2 and 8 and to physically check this sample to choose between codes 1 and 3.

4. Conclusions

There is an increasing need in the mining industry for high performance of mineral resource models. A portion of model deviations are caused by mislogged samples, for which the logged value of a petrophysical attribute such as the lithology, alteration or mineralogical assemblage is erroneous.

Considering the regionalized nature of petrophysical attributes and their dependence relationships with quantitative variables from geochemical analyses or metallurgical tests, a geostatistical approach based on leave-one-out cross-validation has been proposed for identifying possible mislogged samples. The proposal is aimed at calculating, for each sample, a measure of consistency between the logged classes and the quantitative covariates. It is worthwhile to mention that the application is not black-boxed and allows the practitioner to include additional criteria (e.g., *p*-value limits and geographical criteria) to detect suspicious samples. Also, because some samples may have all their measures of consistency smaller than the chosen *p*-value limit and may therefore not be classified into any of the logged classes, the proposed methodology cannot be used blindly as a substitute for the original logs.

To illustrate the applicability of the proposal, a case study from an iron ore deposit has been presented, where the logged rock types are closely related with seven quantitative variables (grades of iron, silica, phosphorus, alumina, manganese, loss on ignition and granulometry) measured on the same set of exploration drill holes. The samples detected as suspicious have been carefully checked on the basis of their logged classes and quantitative covariates, as well as on the basis of the information of the neighboring samples, in order to confirm or reject the correctness of the original rock type logs.

The proposed approach can be applied in several geometallurgical contexts, for example by using geochemical data as covariates for defining lithologies, ore mineralogical data from QEMSCAN (Quantitative Evaluation of Minerals by SCANning electron microscopy) analyses for finding mineral zones, or gangue mineralogical data derived from spectroscopy for recognizing alterations, or all the previous types of data as well as metallurgical tests for identifying geometallurgical domains. In all these contexts, the identification of misslogged samples can be beneficial for the overall performance of the value chain of the mining business and provide criteria to define samples that should be part of a relogging campaign.

Acknowledgments

The authors are grateful to Alejandro Cáceres from University of Chile, as well to the reviewers and editors of the journal, for their constructive comments on a previous version of this work. The funding from the Chilean Commission for Scientific and Technological Research, through Projects CONICYT/FONDECYT/REGULAR/N °1130085 and CONICYT PIA Anillo ACT1407, is also acknowledged.

References

Agterberg, F.P., 1990. Automated Stratigraphic Correlation. Elsevier, Amsterdam, p. 423.
Bourgine, B., Prunier-Leparmentier, A.M., Lembezat, C., Thierry, P., Luquet, C., Robelin, C., 2008. Tools and methods for constructing 3D geological models in the urban environment. The Paris case. In: Ortiz, J.M., Emery, X. (Eds.), Proceedings of the Eighth International Geostatistics Congress. Gecamin Ltda, Santiago, pp. 951–960.
Bourgine, B., Lasseur, E., Leynet, A., Badinier, G., Ortega, C., Issautier, B., Bouchet, V., 2015. Building a geological reference platform using sequence stratigraphy combined with geostatistical tools. In: Geophysical Research Abstracts, 17. EGU2015-8292.
Cáceres, A., Emery, X., 2013. Geostatistical validation of geological logging. In: Ambrus, J., Beniscelli, J., Brunner, F., Cabello, J., Ibarra, F. (Eds.), Proceedings of the Third International Seminar on Geology for the Mining Industry. Gecamin Ltda, Santiago, pp. 73–80.
Chilès, J.P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, p. 699.
Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, p. 369.
Deutsch, J.L., Palmer, K., Deutsch, C.V., Szymanski, J., Etsell, T.H., 2016. Spatial modeling of geometallurgical properties: techniques and a case study. Nat. Resour. Res. 25, 161–181.
Dorr, J.V.N., 1964. Supergene iron ores of Minas Gerais, Brazil. Econ. Geol. 59 (7), 1203–1240.
Dunn, K.J., Bergman, D.J., Latorraca, G.A., 2002. Nuclear Magnetic Resonance – Petrophysical and Logging Applications. Pergamon, Amsterdam.
Emery, X., 2010. Iterative algorithms for fitting a linear model of coregionalization. Comput. Geosci. 36 (9), 1150–1160.
Ewusi, A., Kuma, J.S., 2011. Calibration of shallow borehole drilling sites using the electrical resistivity imaging technique in the granitoids of Central Region, Ghana. Nat. Resour. Res. 20, 57–63.
Glacken, I.M., Snowden, D.V., 2001. Mineral resource estimation. In: Edwards, A.C. (Ed.), Mineral Resource and Ore Reserve Estimation – The AusIMM Guide to Good Practice. The Australasian Institute of Mining and Metallurgy, Melbourne, pp. 189–198.
Haldar, S.K., 2013. Mineral Exploration: Principles and Applications. Elsevier, Oxford, p. 334.
Hassibi, M., Ershaghi, I., Aminzadeh, F., 2003. High resolution reservoir heterogeneity characterization using recognition technology. Dev. Petrol. Sci. 51, 289–307.
Hearst, J.R., Nelson, P.H., Paillet, F.L., 2000. Well Logging for Physical Properties. Wiley, Chichester.
Hoyle, I.B., 1986. Computer techniques for the zoning and correlation of well-logs. Geophys. Prospect. 34 (5), 648–664.
Knödel, K., Lange, G., Voigt, H.J., 2007. Environmental Geology: Handbook of Field Methods and Case Studies. Springer, Berlin, p. 1357.
Luthi, S.M., 2001. Geological Well Logs: Their Use in Reservoir Modeling. Springer, Berlin, p. 373.
Luthi, S.M., Bryant, I.D., 1997. Well-log correlation using a back-propagation neural network. Math. Geol. 29 (3), 413–425.
Maleki, M., Emery, X., 2015. Joint simulation of grade and rock type in a stratabound copper deposit. Math. Geosci. 47, 471–495.
Manchuk, J.G., Deutsch, C.V., 2012. Applications of data coherency for data analysis and geological zonation. In: Abrahamsen, P., Hauge, R., Kolbjørnsen, O. (Eds.), Geostatistics Oslo 2012. Springer, pp. 173–184.
Marjoribanks, R., 2010. Geological Methods in Mineral Exploration and Mining. Springer, Berlin, p. 238.
Moon, C.J., Whateley, M.K.G., Evans, A.M., 2006. Introduction to Mineral Exploration. Blackwell Scientific Publications, Oxford, p. 481.
Rossi, M.E., Deutsch, C.V., 2014. Mineral Resource Estimation. Springer, Heidelberg, p. 332.
Sinclair, A.J., Blackwell, G.H., 2002. Applied Mineral Inventory Estimation. Cambridge University Press, Cambridge, p. 400.
Soleimani, M., Shokri, B.H., Rafiei, M., 2016. Integrated petrophysical modeling for a strongly heterogeneous and fractured reservoir, Sarvak Formation, SW Iran. Nat. Resour. Res.
Soltani, S., Hezarkhani, A., 2011. Determination of realistic and statistical value of the information gathered from exploratory drilling. Nat. Resour. Res. 20, 207–216.
Spies, B.R., 1996. Electrical and electromagnetic borehole measurements: a review. Surv. Geophys. 17, 517–556.
Taylor, G.R., 2000. Mineral and lithology mapping of drill core pulps using visible and infrared spectrometry. Nat. Resour. Res. 9, 257–268.
Theys, P., 1999. Log Data Acquisition and Quality Control. Editions Technip, Paris, p. 480.
Verly, G., 1983. The multigaussian approach and its application to the estimation of local reserves. Math. Geol. 15 (2), 259–286.
Wackernagel, H., 2003. Multivariate Geostatistics: An Introduction With Applications. Springer, Berlin, p. 387.
Yunsel, T.Y., Ersoy, A., 2011. Geological modeling of gold deposit based on grade domaining using plurigaussian simulation technique. Nat. Resour. Res. 20, 231–249.