

MAF DECOMPOSITION OF COMPOSITIONAL DATA TO ESTIMATE GRADES IN IRON ORE DEPOSITS

Maria Noel Morales Boezio ¹, Luciana Arnt Abichequer ² and João Felipe Coimbra Leite Costa³

Abstract Products from iron ore deposits are defined by iron grade, contaminants grades and grain sizes. Furthermore, the data sets constitute regionalized compositions, in which each part carries information that is relative to a total, provided both by the mass balances for each attribute among granulometric partitions, and in each granulometric partition among attributes of interest. In this context, transformation through the additive log-ratio transform (*alr*) makes it possible to use the classical approach of ordinary cokriging, and then back-transform them into the simplex - the sample space of compositional data where the constant sum constraint is maintained. However, implementing cokriging for multiple variables has the known disadvantage of modeling the correogionalization, where the difficulties in satisfying the positive definiteness conditions and the lack of adherence of the models, increases with the number of variables. An approach to override the difficulties related to modeling the LMC is to decompose the multiple correlated random functions through orthogonal factors with no spatial correlation. Thus, there is no need of modeling the correogionalization: each factor can be independently modeled and treated as an independent variable. The decomposition through Min/Max Autocorrelation Factors (MAF) has the advantage, when compared with the classical decomposition in principal components, of decorrelating variables for separation vectors different from zero. In this work, decomposition through MAF is implemented over the additive log-ratio transformations, in order to estimate multiple variables that constitute a regionalized composition, with results that satisfy the original considered balances and that provide adequate results (reproduction of global and local mean, no negative values within the data values interval). Results obtained combining both the compositional data approach and MAF decomposition, proved to better when compared to the ones obtained through cokriging of raw data. The MAF decomposition, in addition, eases the computational and operational efforts of modeling the correogionalization.

¹ Laboratory of Mineral Exploration, Mining Engineering Department, Federal University of Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, Bloco IV, Prédio 75, Sala 104, CEP 91501-970, Porto Alegre, Brazil, maria.morales@ufrgs.br

² Laboratory of Mineral Exploration, Mining Engineering Department, Federal University of Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, Bloco IV, Prédio 75, Sala 104, CEP 91501-970, Porto Alegre, Brazil, luabichequer@hotmail.com

³ Laboratory of Mineral Exploration, Mining Engineering Department, Federal University of Rio Grande do Sul (UFRGS), Av. Bento Gonçalves 9500, Bloco IV, Prédio 75, Sala 104, CEP 91501-970, Porto Alegre, Brazil, jfelipe@ufrgs.br

Introduction

The process of geological research that assists the determination of grades and tonnages of a mineral deposit includes analyzing many variables, either because of their economic value to the assessment of a subsequent ore processing operation, or to understand the geological processes of formation. The joint consideration of correlated variables for grades estimation is much more consistent with phenomenon under study.

Multivariate geostatistics allows the utilization of classical statistical/geostatistical techniques to estimate multiple corregeionalized variables [1] that will assist in the ore body characterization or to estimate a single variable using the information from others as secondary information. Among the multivariate geostatistics methodologies, cokriging [2] is the most classic one providing an unbiased estimator which minimizes the error variance [3]. Its major drawback is that the spatial variability of correlated data must first be jointly modeled to be used in the regression models required.

Many models of corregeionalization have emerged to solve the problem of modeling joint spatial variability, being the Linear Model of Corregeionalization (LMC) the most widespread, where all direct and cross variograms are linear combinations of the same basic structures. The Intrinsic Model of Corregeionalization (IMC) is a particular case in which all variograms are proportional to a unique model [3].

Modeling of the corregeionalization is the major drawback of cokriging. It makes the procedure discouraging to be used by the mining industry, which requires consistent results but fast retrieval [4, 5, 6, 7]. An approach to override these difficulties is to decompose the multiple correlated random functions through orthogonal factors with no spatial correlation. Decomposition through MAF (minimum/maximum autocorrelation factors) [8, 9, 10] decorrelates regionalized variables up to a small separation vector, allowing them to be estimated independently.

In parallel with the problem stated above, there are data sets that constitute regionalized compositions, in which each part carries information that is relative to a total. Examples of this situation occur in bauxite, phosphates, manganese and iron ores, where the regionalized compositions are provided both by the mass balances for each attribute among granulometric partitions, and in each granulometric partition among attributes of interest. In these cases, the correlation might be spurious, given by the closed sum (mass balances) [11] leading to negative estimates due to the negative bias condition [12, 13]. In addition, there is also a concern regarding the reproduction of the closed sum by the estimates.

In this context, transformation of simplicial coordinates into the real space through the additive log-ratio transform (*alr*) makes it possible to use the classical approach of ordinary cokriging, and then back-transform them into their sample space, the simplex [14]. Although isometric log-ratio transformation (*ilr*) is mathematically better than additive log-ratio (*alr*), because it preserves the metric of the simplex [13], the ease of use of the additive log-ratio (*alr*) makes it a better choice for everyday application in the industry.

Due to the mentioned difficulties when modeling the spatial correlogram, this work proposes a methodological approach consisting in decorrelation of the additive log-ratios (*alr*) through MAF decomposition, estimating each factor independently and successively back transforming the estimated factors into the *alr* real space and then into the simplex to satisfy the closed sum. This approach is analogous to the non-regionalized case of principal components factorization of compositional data [15]. The results are compared both to the ones obtained through cokriging of the original variables and through cokriging of the additive log-ratios (*alr*).

Methodology

The methodologies used in the present work are ordinary cokriging, MAF decomposition and transformation in additive log-ratios (*alr*). Thorough revisions of these methodologies can be found in [16], [9] and [12], respectively.

Case Study

The case study comes from a BIF (banded iron formation) iron ore deposit, located in the Ferrous Quadrilateral, Brazil. The data set comes from various types of itabirites with economic importance, with iron grades ranging from 30 to 64% that constitute a geostatistical domain, arbitrarily called IB.

There are three iron ore products given by the grain size partition: Lump Ore, Sinter Feed and Pellet Feed. Dimensioning of the processing plant makes it necessary to sub-divide the Sinter Feed fraction in two. One of these sub-fractions will be referred to as “21D” and is the one chosen for the case study.

Analysis in each granulometric partition lead to the attributes of interest: grain size partition, iron, alumina, silica, phosphorous, manganese and loss on ignition, namely W_i , FE_i , AL_i , SI_i , PI_i , MN_i and PPC_i , respectively, where index i , in this case, corresponds to granulometric partition “21D”.

Mass balance that leads to the constant sum is given by Equation (1).

$$\frac{FE_{21D}(u)}{0.69825} + \frac{P_{21D}(u)}{0.43638} + \frac{MN_{21D}(u)}{0.63193} + AL_{21D}(u) + SI_{21D}(u) + PPC_{21D}(u) = 100 \quad (1)$$

Closure operation

The original values in the data set do not add up to 100% as stated in Equation (1). For this reason, the closure operation is performed, following Equation (2).

$$Z_clsd(u) = \left(\frac{Z(u)}{\text{sum}(21D)} \right) \cdot 100 \quad (2)$$

where $Z_clsd(u)$ is the new variable after the closure operation,
 $Z(u)$ is the original value in the data set,
and $\text{sum}(21D)$ is the value that corresponds to the addition of the original values in fraction 21D.

Table 1 presents the basic statistics of the original data after closure, including declustered mean and variance [17] obtained with a moving window [17] of dimensions 200 x 100 x 10m in directions X, Y and Z.

Table 1: Basic statistics of the original (closed) data, including declustered mean and variance.

	Num. of Samples	Min	Max	Mean	Var	Declustered Mean	Declustered Var
<i>AL21D_clsd</i>	909	0.10	3.11	0.79	0.23	0.79	0.24
<i>FE21D_clsd</i>	909	21.53	68.78	60.96	39.19	60.84	43.03
<i>MN21D_clsd</i>	909	0.01	9.90	0.30	0.59	0.30	0.49
<i>P21D_clsd</i>	909	0.01	0.27	0.06	0.002	0.06	0.002
<i>PPC21D_clsd</i>	909	0.04	9.76	2.54	3.20	2.54	3.39
<i>SI21D_clsd</i>	909	0.64	68.39	8.89	78.15	8.92	89.11

Scatter plots of original values versus new closed ones are presented in Figure 1 to evaluate the influence of the closure operation. The value of the closed variables is identical to that of the original ones, except in the case of iron grade. Nevertheless, the correlation coefficient in this case is almost 1 (one).

Additive log-ratio (*alr*) transformation

The constant sum constraint is given by Equation (1). For obtaining the additive log-ratios (*alr*) $Y(u)$, silica is arbitrarily chosen to be the divisor in the ratios presented in Equations (3) to (7).

$$Y_1(u) = \ln \left(\frac{FE21D_clsd(u)}{0.69825 \cdot SI21D_clsd(u)} \right) \quad (3)$$

$$Y_2(u) = \ln \left(\frac{P21D_clsd(u)}{0.43683 \cdot SI21D_clsd(u)} \right) \quad (4)$$

$$Y_3(u) = \ln\left(\frac{MN21D_clsd(u)}{0.63193 \cdot SI21D_clsd(u)}\right) \quad (5)$$

$$Y_4(u) = \ln\left(\frac{AL21D_clsd(u)}{SI21D_clsd(u)}\right) \quad (6)$$

$$Y_5(u) = \ln\left(\frac{PPC21D_clsd(u)}{SI21D_clsd(u)}\right) \quad (7)$$

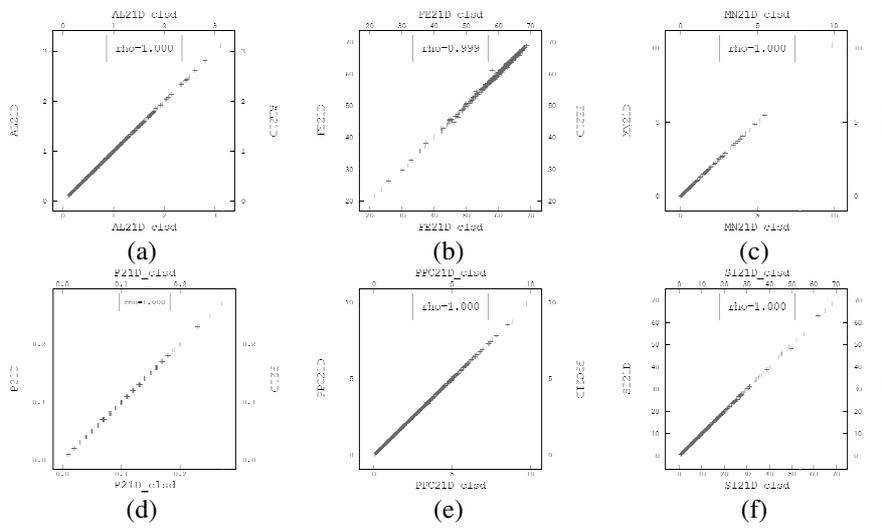


Figure 1: Scatter plots of closed (_clsd) vs. original data with rho (correlation coefficient) for variables alumina (a), iron (b), manganese (c), phosphorous (d), loss on ignition (e) and silica (f).

As the *alr* transformation removes the spurious correlation, but does not guarantee the spatial decorrelation of the resulting transformed variables, spatial continuity of the additive log-ratios leads to an ellipsoid of anisotropy with its major and intermediate axes respectively aligned along directions N90, and N0° in the XY plane, and minor one perpendicular to this plane. Corregionalization is modeled through the Linear Model of Corregionalization (LMC) [16, 3] with a nugget effect C_0 and two spherical structures *Sph*, with contributions to global variance C_1 and C_2 , as shown in Equation (8).

$$\gamma(h) = C_0 + C_1 \cdot Sph\left(\frac{365m}{N90^\circ} \frac{280m}{N180^\circ} \frac{90m}{D90^\circ}\right) + C_2 \cdot Sph\left(\frac{2100m}{N90^\circ} \frac{400m}{N180^\circ} \frac{150m}{D90^\circ}\right) \quad (8)$$

Afterwards, ordinary cokriging is performed in the additive log-ratios (*alr*), and the estimates are back transformed to the simplex through Equations (9) to (15).

$$FE21_ckALR(u) = 0.69825 \cdot 100 \cdot \exp(ALR1^*(u)) / K(u) \quad (9)$$

$$P21_ckALR(u) = 0.43683 \cdot 100 \cdot \exp(ALR2^*(u)) / K(u) \quad (10)$$

$$MN21_ckALR(u) = 0.63193 \cdot 100 \cdot \exp(ALR3^*(u)) / K(u) \quad (11)$$

$$AL21_ckALR(u) = 100 \cdot \exp(ALR4^*(u)) / K(u) \quad (12)$$

$$PPC21_ckALR(u) = 100 \cdot \exp(ALR5^*(u)) / K(u) \quad (13)$$

$$SI21_ckALR(u) = 100 / K(u) \quad (14)$$

$$K(u) = 1 + \exp(ALR1^*(u)) + \exp(ALR2^*(u)) + \exp(ALR3^*(u)) + \exp(ALR4^*(u)) + \exp(ALR5^*(u)) \quad (15)$$

where $ALR1^*(u)$, $ALR2^*(u)$, $ALR3^*(u)$, $ALR4^*(u)$ and $ALR5^*(u)$, are the additive log-ratios (*alr*) estimates obtained through ordinary cokriging, and $FE21_ckALR(u)$, $P21_ckALR(u)$, $MN21_ckALR(u)$, $AL21_ckALR(u)$, $PPC21_ckALR(u)$ and $SI21_ckALR(u)$ are the estimates of the variables of interest after back transformation into the simplex.

MAF decomposition

MAF decomposition consists in transforming a multivariate random vector into a set of orthogonal factors, extending the principal components approach [18] up to a separation vector (*h*) greater than zero, generally coincident with the spacing among samples or the range of the first structure of the LMC [9].

MAF decomposition is performed choosing a separation vector $h=150$ meters. Decorrelation among variables is visually verified analyzing the cross-variograms of the factors, where it is observed an almost pure nugget effect for the separation vectors up to the selected distance and a decreasing autocorrelation from the first factor MAF1_ALR (maximum) to the fifth one MAF5_ALR (minimum).

Spatial variability of each factor is analyzed and modeled independently with a nugget effect and two spherical structures (Figure 2), with the same anisotropy directions that in the *alr* transformation case: principal and intermediate axes along $N90^\circ$ and $N0^\circ$ in the XY plane, and the minor perpendicular to this plane.

Afterwards, each factor is estimated individually through ordinary kriging [1], with the same neighborhood and search strategies that in the prior case of cokriging of the additive log-ratios (*alr*).

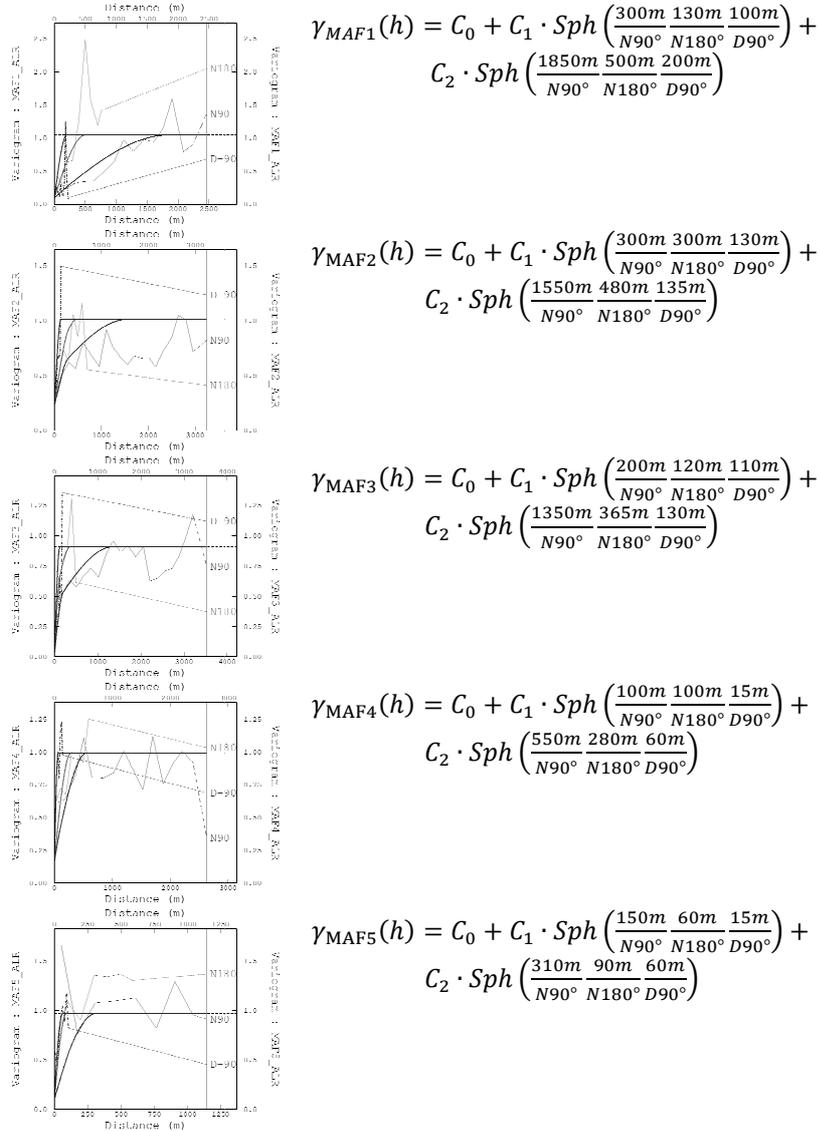


Figure 2 – Variograms of MAF factors, showing decreasing spatial correlation from MAF1 (maximum) to MAF 5 (minimum).

Afterwards, the estimates are back transformed into additive log-ratios (*alr*), and again into the simplex through Equations (9) to (15), as before. The estimates

in this case are $FE21_bckMAF(u)$, $P21_bckMAF(u)$, $MN21_bckMAF(u)$, $AL21_bckMAF(u)$, $PPC21_bckMAF(u)$ and $SI21_bckMAF(u)$.

Cokriging approach

Cokriging of the original (closed) variables is performed with a Linear Model of Coregionalization (LMC) to model the joint spatial correlation. In this case, the ellipsoid of anisotropy is slightly rotated, with its principal and intermediate axes along directions N100° and N190° respectively in the XY plane, and the minor one is perpendicular to this plane.

In this case, LMC is more difficult to model, as there are six variables involved. It is modeled, as in the previous cases, with a nugget effect, C_0 and two spherical structures Sph , with contributions to global variance C_1 and C_2 , as it is stated in Equation (16).

$$\gamma(h) = C_0 + C_1 \cdot Sph\left(\frac{365m \ 280m \ 90m}{N100^\circ \ N190^\circ \ D90^\circ}\right) + C_2 \cdot Sph\left(\frac{1245m \ 450m \ 150m}{N100^\circ \ N190^\circ \ D90^\circ}\right) \quad (16)$$

Ordinary cokriging is then performed using the same neighborhood and search strategies parameters than in the cases presented before. Estimates in this case have the same notation that the original closed data: $FE21D_clsd(u)$, $P21D_clsd(u)$, $MN21D_clsd(u)$, $AL21D_clsd(u)$, $PPC21D_clsd(u)$ and $SI21D_clsd(u)$.

Discussion of results

One of the aspects discussed next is the unbiasedness of the estimates, analyzed through the reproduction of the global and local mean. Reproduction of the global mean can be observed in Table 2. This table also shows that in the case of cokriging of the original (closed) data, there are estimates outside the original data interval, and some negative values as in the case of manganese.

For checking the reproduction of the local mean, the conditional expectation of the estimates and original data are plotted along principal directions X, Y and Z using swath plots. In this paper, only iron, alumina and phosphorous grades along X direction are shown (Figure 3).

From these diagrams it can be stated that the local mean of the data set is reproduced by the estimates. Moreover, when performing cokriging of the additive log-ratios (alr) and their MAF decomposition with independent kriging, results are very similar. For this reason, scatter-plots of the estimates obtained by the different methodologies are presented in Figure 4.

Correlation coefficients (ρ) exceed 0.90, indicating that the estimates are very similar, especially in the cases where the alr transformation is performed.

For the iron estimates, there is a slight tendency of the methods that involve additive log-ratio (*alr*) transformation, to provide higher values than direct cokriging of the original (closed) data, as is noticed in the swath and scatter plots. Although cross validation [17] was implemented for all the variables, variograms, search neighborhood and strategies and that Table 2 shows a good reproduction of the global mean, further modifications in the kriging/cokriging parameters should be implemented to obtain better results.

Table 2 – Statistics of original (closed) data (**_clsd** in bold), estimates obtained through ordinary cokriging (*_clsd*), cokriging of additive log-ratios *alr* (*_ckALR*) and kriging of MAFs obtained from additive log-ratios *alr* (*_bckMAF*). (*) denotes values that are outside the original data interval.

	Count	Minimum	Maximum	Mean
AL21D_clsd	909	0.10	3.11	0.79
AL21D_clsd	20813	0.14	2.19	0.88
AL21_ckALR	20823	0.13	2.11	0.82
AL21_bckMAF	20823	0.11	2.25	0.83
FE21D_clsd	909	21.53	68.78	60.84
FE21D_clsd	20813	37.39	68.85*	61.35
FE21_ckALR	20823	26.58	68.01	62.50
FE21_bckMAF	20823	26.67	67.91	62.53
MN21D_clsd	909	0.01	9.90	0.30
MN21D_clsd	20813	-0.09*	2.40	0.27
MN21_ckALR	20823	0.01	3.25	0.14
MN21_bckMAF	20823	0.01	4.35	0.14
P21D_clsd	909	0.01	0.27	0.06
P21D_clsd	20813	0.00*	0.22	0.06
P21_ckALR	20823	0.01	0.15	0.06
P21_bckMAF	20823	0.01	0.18	0.06
PPC21D_clsd	909	0.04	9.76	2.54
PPC21D_clsd	20813	0.03*	8.46	2.79
PPC21_ckALR	20823	0.08	8.28	2.62
PPC21_bckMAF	20823	0.09	7.71	2.63
SI21D_clsd	909	0.64	68.39	8.92
SI21D_clsd	20813	0.01*	45.62	7.89
SI21_ckALR	20823	1.52	61.08	6.70
SI21_bckMAF	20823	1.49	60.95	6.62

Another aspect to be analyzed is the closed sums.

The constant sums are perfectly satisfied for the totality of the estimates in the cases of cokriging of the additive log-ratios *alr* and also when kriging independently the MAF obtained from the additive log-ratios *alr* (Table 3), with estimates that are in the original sample space where the constant sum is maintained (the simplex).

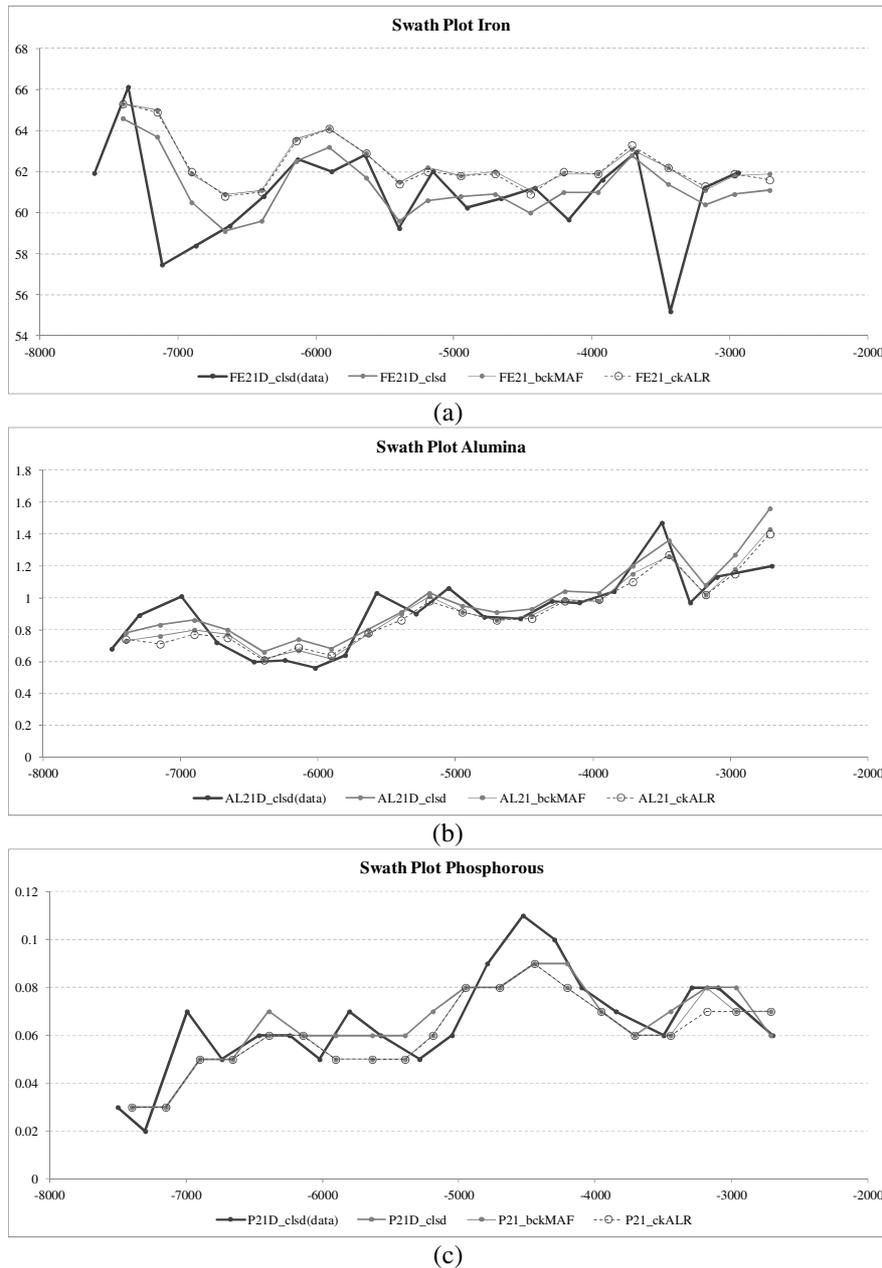


Figure 3 – Swath plots: conditional expectation of the estimates and original, plotted together along direction X, for iron (a), alumina (b) and phosphorous (c) grades.

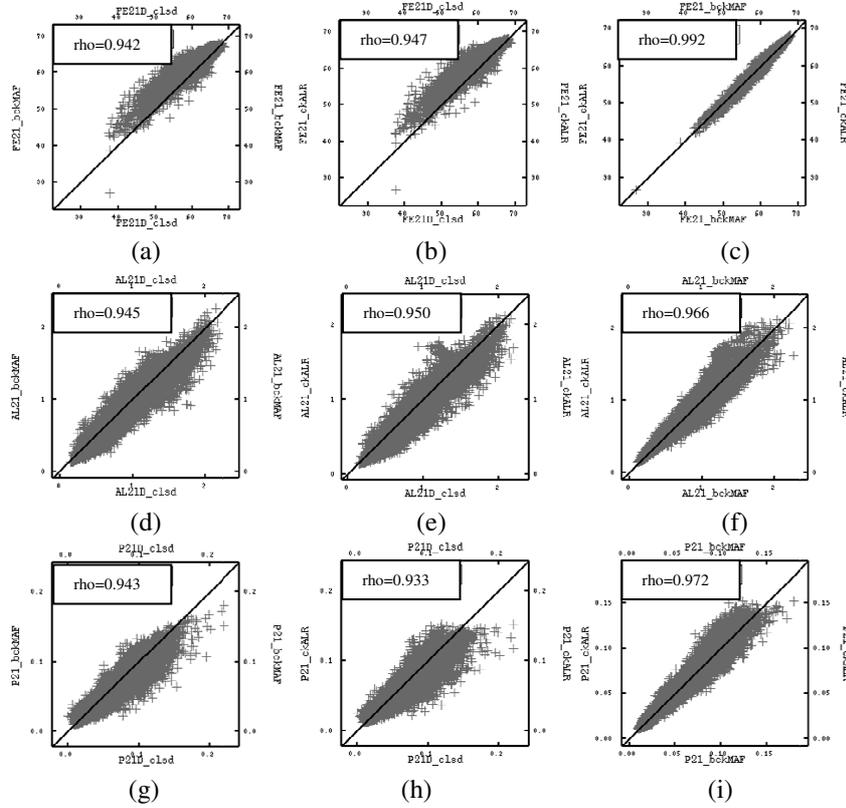


Figure 4 – Scatter plots of the estimates obtained through kriging of the MAF vs. cokriging of the original (closed) data, cokriging of *alr* vs. cokriging of the original (closed data) and kriging of MAF vs. cokriging of *alr*, for iron ((a) (b) and (c)), alumina ((d), (e) and (f)) and phosphorous ((g), (h) and (i)) grades, respectively.

Table 3 – Statistics of the sum of the estimates in each block, obtained through cokriging of the original (closed) data (*_clsd*), cokriging of the additive log-ratios *alr* (*_ckALR*) and kriging of the MAF obtained from the additive log-ratios (*_bckMAF*).

	Count	Minimum	Maximum	Mean	Variance
<i>sum_clsd</i>	20823	99.791	100.230	100.000	0.001
<i>sum_ckALR</i>	20823	100.000	100.000	100.000	0.000
<i>sum_bckMAF</i>	20823	100.000	100.000	100.000	0.000

The constant sums obtained through the estimates obtained by ordinary cokriging of the original data ordinary are adequate, but this situation cannot be extrapolated to another data set, because ordinary cokriging of the original (closed) data does not guarantee that the constant sum is satisfied.

Conclusions

Decomposition through MAF was implemented over additive log-ratio transformations *alr* with the objective of estimating multiple variables that constitute a regionalized composition. This is the case of variables from iron ore and other deposits such as manganese, bauxite and phosphates.

Apart from providing unbiased estimates which reproduce the global and local mean, this combination of methodologies aims to solve two drawbacks when performing classical ordinary cokriging to improve the results by utilization of the joint spatial correlation:

(i) reproduction of the original closed sums from the mass balances among chemical species and granulometric partitions,

(ii) simplification of the computational and operational efforts of modeling the corregionalization through the Linear Model of Corregionalization (LMC) or alternatively, the Intrinsic Model of Corregionalization (IMC).

Results obtained through cokriging of the additive log-ratios (*alr*) provided non-negative estimates, within the original data intervals, with an adequate reproduction of global and local mean, with the totality of the estimates adding up to the desired constant of 100%.

When decomposing the additive log-ratios (*alr*) through MAF, the results were almost identical to the ones obtained through cokriging of the *alr*, through a much easier and simpler implementation.

For the reasons mentioned above, it can be said that combination of both methodologies provided better results than cokriging of raw data – unbiased positive estimates, within the original data interval, adding up to 100% - without the effort of modeling the corregionalization.

Bibliography

[1]. **Matheron, G.** *Les variables regionalisées et leur estimation*. Paris : Ed. Masson, 1965.

[2]. **Marechal, A.** *Cokrigage et Regression em Correlation Intrinsique*. Fontainebleau : Centre de Geostatistique de Fontainebleau, 1970.

[3]. **Wackernagel, H.** *Multivariate Geostatistics. An Introduction with Applications*. Berlin : Springer-Verlag, 1994.

[4]. *Geoestatística multivariada aplicada a depósitos de ferro - Complexo Pico - MG.* **Boezio, M. N. M., Costa, J. F. C. L., Ribeiro, D. T., Pilger, G. G. and Koppe, J. C.** Belo Horizonte, Brazil : 4th Brazilian Congress on Open Pit and Underground Mining, IBRAM, 2006.

[5]. *Decomposição MAF como alternativa à cokrigagem na modelagem de depósitos de ferro - Mina Sapecado - MG.* **Boezio, M. N. M., Costa, J. F. C. L. and Koppe, J. C.** Belo Horizonte, Brazil : 5th Brazilian Congress on Open Pit and Underground Mining, IBRAM, 2008.

- [6]. **Boezio, M. N. M.** *Estudo das Metodologias Alternativas da Geoestatística Multivariada Aplicada a Estimativa de Teores de Depósitos de Ferro*, DEng Thesis, Tutor: Costa, Joao Felipe Coimbra Leite. Porto Alegre, RS, Brazil : Federal University of Rio Grande do Sul, UFRGS, 2010.
- [7]. *Multivariate Geostatistics in Complex Iron Ore Deposits*. **Boezio, M. N. M., Costa, J. F. C. L. and Koppe, J. C.** Santiago, Chile : GEOSTATS 2008 Proceedings of the Eight International Geostatistics Congress, Ed. Ortiz, J. M. and Emery, X., 2008.
- [8]. **Switzer, P. and Green, A. A.** *Min/Max autocorrelation factors for multivariate spatial imagery*. Standford : Standford University Press, 1984.
- [9]. *Geostatistical simulation of regionalized pore-size distributions using Min/Max autocorrelation factors*. **Desbarats, A.J. and Dimitrakopoulos, R.** 2000, Mathematical Geology, pp. 919-942.
- [10]. **Boucher, A.** *Conditional Joint Simulation of Random Fields on Block Support*, MSc Thesis. Tutor, Dimitrakopoulos, R. Brisbane, Australia : University of Queensland, 2003.
- [11]. **Aitchison, J.** *The Statistical Analysis of Compositional Data, Monographs on Statistics and Applied Probability*. s.l. : Chapman & Hall Ltda, 1986.
- [12]. **Pawłowsky, V. and Olea, R. A.** *Geostatistical Analysis of Compositional Data, Studies in Mathematical Geology*. s.l. : Oxford university Press, 2004.
- [13]. **Tolosana-Delgado, R.** *Geostatistics for constrained variables: positive data, compositions and probabilities. Application to environmental hazard monitoring [PhD thesis]*. Girona : University of Girona, 2006.
- [14]. **Mateu-Figueiras, G.** *Models de distribució sobre el simplex [PhD thesis]*. Barcelona : Universidade Politècnica de Catalunya, 2004.
- [15]. *Principal components analysis of compositional data*. **Aitchison, J.** 1983, Biometrika, pp. 57-65. 1.
- [16]. **Chliés, B.M. and Delfiner, P.** *Geostatistics: modeling spatial uncertainty*. New York : Wiley-Interscience, 1999.
- [17]. **Isaaks, E.H and Srivastava, R.M.** *An Introduction to Applied Geostatistics*. New York : Oxford University Press, 1989.
- [18]. **Johnson, R.A. and Wichern, D.W.** *Applied Multivariate Statistical Analysis*. New Jersey : Prentice Hall, 1998.