Imputation of Reactive Silica and Available Alumina in Bauxites by Self-Organizing Maps

Cleyton de Carvalho Carneiro[1], Dayana Niazabeth Del Valle Silva Yanez[2], Carina Ulsen[1], Stephen J. Fraser[3], Juliana Lívi Antoniassi[1], Simone P. A. Paz[4], Rômulo Simões Angélica[5], Henrique Kahn[1]

[1]Universidade de São Paulo, Escola Politécnica, Dpto. de Engenharia de Minas e de Petróleo;
[2]Universidad Simón Bolívar, Dpto. Ciencias de la Tierra;
[3]CSIRO Mineral Resources (Mining Systems), Queensland Centre for Advanced Technologies;
[4]Universidade Federal do Pará, Faculdade de Engenharia de Materiais;
[5]Universidade Federal do Pará, Instituto de Geociências, Laboratório de Caracterização Mineral.

Corresponding author: cleytoncarneiro@usp.br - *Dpto. de Engenharia de Minas e de Petróleo, Escola Politécnica da Universidade de São Paulo - Av. Prof. Mello Moraes, 2373, Butantã - CEP 05508-030, São Paulo – SP, Brasil*

***Abstract* - Geochemical analyses can provide multiple analytical variables. Accordingly, the generation of large geochemical databases enables imputation studies or analytical estimates of missing values or complex measuring. The processing of bauxite is a key step in the production of aluminum, in which the determination of Reactive Silica ($RxSiO_2$) and Available Alumina ($AvAl_2O_3$) are very relevant. The traditional analytical method for achieving $RxSiO_2$ has limitations associated with poor repeatability and reproducibility of results. Based on the values from the unsupervised Self-Organizing Maps technique, this study aims to develop, systematically, the imputation of missing grades of the geochemical composition of bauxite samples of a database from three trial projects, for the variables: total Al_2O_3; total SiO_2; total Fe_2O_3; and total TiO_2. Each project was submitted to partial exclusion of $AvAl_2O_3$ and $RxSiO_2$ values, in proportion of 20%, 30%, 40% and 50%, to investigate the SOM technique as imputation method for $RxSiO_2$ and $AvAl_2O_3$. By comparing the imputed values from the SOM analysis with the original values, SOM technique demonstrated to be an imputation tool capable of obtaining analytical results with up to 50% of missing data. Specifically, the best results demonstrate that $AvAl_2O_3$ can be obtained by imputation with a higher correlation than $RxSiO_2$, based on the parameters and variables involved in the study. Similarity in the nature of samples and an increase in the number of embedded analytical variables are factors that provided better imputation results.**

Index Terms* - *Analytical Geochemical Imputation, Self-Organizing Maps (SOM), Bauxite, Reactive Silica, Available Alumina.

I. Introduction

Currently, in the fields of geophysics and geochemistry, there are numerous advances that allow the acquisition of a high density, of multivariate samples. The analysis of these data, however, requires greater methodological studies, especially regarding the optimization of relationships between different variables. During the exploratory phase of database analysis, it is possible to improve the integration and interpretation, as well as to impute and / or estimate analytical values.

A data imputation approach can be developed from statistical methods (univariate, causal or multivariate), or even by intelligent and non-statistical methods (such as Fuzzy Logic, Supervised Artificial Neural Networks, Self-Organizing Maps and Genetic Algorithm). However, in relation to the non-linear and non-parametric series, Self-Organizing Maps (SOM) presents some advantages, since the method is based on the organization in an n-dimensional space, whose results are projected as a map, preserving the topological relations of similarity.

According to [1] SOM can be considered an exploratory data analysis tool, and the method can be used to carry out broad categories of operations, such as forecasting or estimating, clustering, classification, and/or noise reduction. Although several imputation methods that can synthetically create values in areas with missing data, Self-Organizing Maps (SOM) outstands. While most of methods are statistical, and processed in a single variable - such as neighborhood-based interpolations, SOM is a non-linear method, based on the principles of vector quantization and measurement of vector similarity [2].

In a multivariate database, SOM analysis treats each sample as a vector unit. After creating an n-dimensional space, bring “n” is the number of variables involved, the samples elect a Best Matched Unit (BMU), which provides the corresponding values for each variable. It makes possible the creation of synthetic values of samples with missing values from their respective BMUs, in a multivariate database, as proven by [3] and [4].

Therefore, it becomes possible to implement SOM analysis as an alternative tool to impute analytical data. To address such a proposal, quantitative measurement of Available Alumina ($AvAl_2O_3$) content, mostly related to gibbsite, and Reactive Silica ($RxSiO_2$), mainly present in kaolinite, in bauxite deposits from various regions of Brazil were explored.

Initially, XRF chemical analyses were carried out at Technological Characterization Laboratory (LCT) from the Department of Mining and Petroleum Engineering (PMI) of the Polytechnic School of USP [5], $AvAl_2O_3$ and $RxSiO_2$ at external laboratory. Among the variables obtained from the chemical analysis of bauxite samples, the $RxSiO_2$ results showed low reproducibility, repetitiveness and incurred analytical high costs [6], besides dealing with minucious and time-consuming procedures. Thus, imputation techniques, such

978-1-5090-6638-4/17/$31.00 ©2017 IEEE

as SOM, become important by estimating values in the multivariate analytical data.

Bauxitic ore is generally composed of hydroxide or oxyhydroxide aluminum minerals (e.g., gibbsite, boehmite and diaspore) iron and other minor minerals (e.g., mainly kaolinite, hematite, goethite, quartz, anatase), produced by weathering of aluminosilicate rocks under tropical and subtropical climate conditions, typically under high percolation rate hydrological conditions [7], [8], [9], [10]. In Brazil's economic scenario, bauxite has a prominent position, as the country holds the world's third largest reserves and ranks third as a country producer [11].

An extensive database was created by analyzing many bauxites and mineral separation products from the determination of $AvAl_2O_3$ and $RxSiO_2$ content by wet chemistry and others main elements by XRF. The database contains necessary and significant elements to impute unknown values of these variables in other samples. From the multivariate data analysis technique SOM, this research aims to impute unknown values of $AvAl_2O_3$ and $RxSiO_2$ in bauxite samples by BMU from self-organized maps, which represent the relationships of samples in a n-D space of the variables.

The SOM analysis was divided into four phases for each of the three projects A, B and C. The samples were chemically analyzed using the variables: mass recovery (%) and content (wt%) of (i) $AvAl_2O_3$; (ii) $RxSiO_2$, (iii) Total Al_2O_3; (iv) Total SiO_2; (v) Total Fe_2O_3; and (vi) Total TiO_2. The phases consist in the partial omission of 20%, 30%, 40% and 50% analytical values of $AvAl_2O_3$ and $RxSiO_2$ for the samples. The results imputed by the technical SOM were compared with the original analytical geochemical values, and evaluated according to descriptive statistics.

Once demonstrated, the correlation between original data and impute analytics by SOM can assist users of chemical analysis in bauxites in getting $AvAl_2O_3$ and $RxSiO_2$ content with low cost material and labor. The analytical and probabilistic tests on these results would bring safety and reliability for the use of the tool as a reference factor in other analyses.

The experiments developed in this project aim to address challenges of the validation of a new system to achieve $AvAl_2O_3$ and $RxSiO_2$ content, with low running costs without compromising the quality of results standards. Furthermore, the results aimed at promoting the use of SOM art as imputation tool capable of providing satisfactory analytical results. The results precede the application of SOM in geochemical, geophysical essays or in several other areas where similar uncertainties occur or needs with respect to imputation, integration and interpretation of multivariate data.

II. MATERIALS AND METHODS

A. Bauxite sample selection for SOM analysis

The bauxite samples selected for SOM analysis were part of the three projects' databases (A, B and C), composed by different lithologic characteristics. The interaction between different projects was carried out to obtain the best representation and comparability between them. Samples were prior studied for mineral separability at the Technological Characterization Laboratory (LCT) of the University of São Paulo - Brazil (EPUSP).

Estimation and evaluation of the mineralogical composition of the bauxite samples were developed using chemical analyses of total contents of Al_2O_3, SiO_2, Fe_2O_3 and TiO_2 by X-ray fluorescence (XRF) and specific content of $AvAl_2O_3$ and $RxSiO_2$ by wet chemistry.

Effectively, it was prioritized the chemical compositional variation between high and low $RxSiO_2$ and $AvAl_2O_3$ grades. Thus, the variability of the source of the samples, and the different methods to which they have been subjected, such as classification and separation assays of minerals, have enabled an extensive and diverse database to perform the SOM analysis.

The data acquired from the Project A were composed of 690 samples and the variables characterized were: the total content of Al_2O_3, SiO_2, Fe_2O_3. The data used in Project B rely on the content of Al_2O_3, SiO_2, Fe_2O_3 and TiO_2 in 219 samples. The Project C had a group of 70 samples, characterized in total contents of Al_2O_3, SiO_2, Fe_2O_3 and TiO_2 [6].

B. Experimental procedure

Once selected the samples, it was required a pre-processing of these samples to feed the analysis of imputation values. In the sample pre-processing, random values of $AvAl_2O_3$ and $RxSiO_2$ were excluded for later estimative by the SOM analysis. Finally, the imputed values were compared with the original values obtained by lab chemical analysis.

To measure and assess the magnitude of SOM analysis, sampling data tables with random exclusion of $AvAl_2O_3$ and $RxSiO_2$ values were modified. For each project, 20%, 30%, 40% and 50% of the total samples were randomly hidden for the generation of new tables to be used in imputation values. The data table then was introduced in SOM platform from SiroSOM® software.

TABLE I
SAMPLES PREPARATION

	Samples	E20%	E30%	E40%	E50%	V
Project A	690	138	207	276	345	4
Project B	219	44	66	88	122	7
Project C	70	14	21	28	35	6

E = Exclusion; V = variables.

C. Imputation values from SOM analyses

The $AvAl_2O_3$ and $RxSiO_2$ content imputation of bauxites was developed according to the adaptation of a routine proposal [12], where the data estimated by the technique are based on the distances between available vectors [1]. For data to lower spatial resolution, the traditional estimation process is given by substitution, where the values are produced from the vectors of the BMU's. Often, the data sets will result in biased estimates, which make necessary the use of techniques such as nearest neighbor [13].

A hexagonal grid was chosen as the display format; the surface of a hyper toroidal volume was used for the projection neurons or BMU's. To define the resulting self-organized map size, as in (1) where [n] is the number of samples inserted at SOM platform [14]. Thus, a map size was chosen as suitable for this study. After the generation of the self-organized map, images of U-Matrix and Components Plot were produced.

$$\boldsymbol{Size_{SOM}} = 5x\sqrt{[n]} \qquad (1)$$

The Component Plots (CP) allow visualizing and quantifying the contribution of the analyzed variables for each resulting neuron in self-organized map, allowing the verification of the relationship between the responses of the various components. The U-Matrix enabled the classification of data related to vector similarity constructed from these samples.

As a result, the BMU's were obtained for each sample and analyzed variable, as well as in samples with incomplete analysis. The imputation values were therefore determined from the unique BMU's for each sample, reflecting synthetic representative content for samples where these levels were originally unknown.

It is important to mention that, for allocating data, SiroSOM® code works with the combination of two approaches and variations: (i) replacement of missing values by the BMU's values; (ii) the improvement of the values estimated by an iterative process. In (i), the initial SOM is calculated and determines an initial set of replacement values, whereas in (ii) SOM values are recalculated again to replace the values not found in the input data.

D. *Correlation tests and results evaluation*

The comparison of the results presented by classical analysis and from the SOM analysis was performed using descriptive statistics, with the completion of charts and scatter plots that confront the values obtained for each sample studied of continuous random variables: $AvAl_2O_3$ and $RxSiO_2$.

Dispersion measurements were made around the average, such as variance and covariance, to determine the correlation between calculated and measured variables simultaneously, and the correlation coefficient to the purpose of normalizing the covariance range -1 to 1. Equation 2 calculates the correlation.

$$\boldsymbol{Corr[X,Y]} = \frac{\boldsymbol{Cov(X,Y)}}{\sqrt{\boldsymbol{Var(X)Var(Y)}}} \qquad (2)$$

Accordingly, in a scatter plot, the regression line was shown generating a linear correlation and a set of ordered pairs to determine a functional relationship by the minimum squares method.

The correspondence of the values obtained by SOM analysis and chemical lab analysis was measured by the difference between the mean and median levels, followed by the percentage of relative error, according to (3), where T_{SOM} represents the contents obtained by the SOM tool and T_{LCT} the original contents, obtained by laboratory analysis.

$$\boldsymbol{ERP} = \left|\frac{\boldsymbol{T_{SOM}-T_{LCT}}}{\boldsymbol{T_{LCT}}}\right| * \boldsymbol{100} \qquad (3)$$

In addition, mean and median were calculated for each variable to obtain the percentage error of the samples of each project. According to the comparison of values, it is possible to measure the SOM range by assessing the percentage of error and correlation to determine the effectiveness for omitting $AvAl_2O_3$ and $RxSiO_2$ by 20%, 30% 40% and 50%.

III. Results

A. *Self-Organizing Maps analysis*

To obtain new values of $AvAl_2O_3$ and $RxSiO_2$ 12 self-organized maps were produced at every stage of samples exclusion for the three projects. The number of rows and columns were calculated from the desired size map (SizeSOM).

In addition, at the end of each SOM analysis, the quantization error (QE) was calculated, which represents the average distance between each array and its respective BMU, as the map resolution. In the same way, it has been calculated the final topographic error (Te), which simulates the proportion of all data vectors for which the main BMU (first and second) are not adjacent units.

Table II shows the completed and calculated parameters in the SOM analysis initialization step for each design, the four data deletion steps of 20%, 30%, 40% and 50%, represented by E20% S30% E40%, E50% respectively.

TABLE II
INITIALIZATION STAGE

	Initialization									
	Size SOM		E20%		E30%		E40%		E50%	
	Roll	Col	Qe	Te	Qe	Te	Qe	Te	Qe	Te
Project A	10	14	0.165	0.470	0.147	0.539	0.128	0.545	0.110	0.570
Project B	8	9	0.337	0.128	0.338	0.155	0.330	0.146	0.317	0.228
Project C	6	7	0.282	0.114	0.298	0	0.268	0	0.264	0

E = Exclusion; Qe = Quantization error; Te = Topographic error

Then, in the training process, it was selected the kind of neighbor to each Gaussian vector and defined rough and fine data. Those values were set by default as: Initial radius (Ir1), Final radius (Fr1) and Training Length (TL1), calculated by SOM, presented in Table III.

To represent the structure and pattern of the input samples by the similarities, SOM analysis used the distance data and the initial length for each input sample value.

After initialization and training vectors, CP maps were generated to each variable, along with their integration in the Matrix-U view.

TABLE III
TRAINING STAGE

	Training					
	Ir1	Fr1	TL1	Ir2	Fr2	TL2
Project A	18	5	20	5	1	400
Project B	13	4	20	4	1	400
Project C	10	3	20	3	1	400

Ir = Initial radius; Fr = Final radius; TL = Training Length

IV. Imputed values using Self-Organizing Maps

The SOM technique with SiroSOM® software allowed the imputation of 1414 data pairs excluded from $AvAl_2O_3$ and

$RxSiO_2$ contents, key control elements in the aluminum production chain. Furthermore, from the BMU adjustments, new values were obtained for each sample in the input variables.

A. Correlation and evaluation

To evaluate and simplify the visualization of the results obtained by the SOM, the mean (Mt) and median content (Met) of $AvAl_2O_3$ and $RxSiO_2$ of the original data was calculated, as well as those obtained by the BMU generated by SOM analyses.

This procedure was performed for each project and for the different scenarios of data exclusion. The results are shown in Tables IV, V, VI and VII with the purpose of comparing the influence of deleted data proportion in the use of variables and samples of each project. The correlation graphs between the original and imputed data can be seen in Fig 1.

In general, it can be noted that both $AvAl_2O_3$ and $RxSiO_2$ correlation coefficients as all projects is effective, that is, it maintains a positive, indicating right proportionality. The results are coherent, therefore, considering that they address comparison of the same variable.

TABLE IV

Statistical Analyses $AvAl_2O_3$ and $RxSiO_2$. Exclusion 20%

	Project A				Project B				Project C			
	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2
Mt	36.03	36.19	6.42	6.44	34.75	35.10	8.90	8.86	41.96	41.37	9.45	9.36
Met	39.60	40.33	5.36	5.54	36.19	36.40	5.28	6.14	51.45	49.90	4.60	4.52
ER%	1.81		3.25		0.58		14.01		3.11		1.77	
Corr	0.84		0.03		0.99		0.98		1.00		1.00	
RR	0.71		0.00		0.98		0.95		1.00		0.99	

TABLE V

Statistical Analyses $AvAl_2O_3$ and $RxSiO_2$. Exclusion 30%

	Project A				Project B				Project C			
	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2
Mt	36.07	35.76	6.46	6.84	34.13	34.61	9.18	9.17	36.68	36.86	11.80	11.48
Met	39.41	39.05	5.21	5.52	36.62	37.89	5.30	6.10	48.90	49.06	5.00	6.27
ER%	0.92		5.62		3.35		13.11		0.33		20.26	
Corr	0.78		-0.10		0.99		0.98		1.00		1.00	
RR	0.61		0.01		0.98		0.96		1.00		0.99	

TABLE VI

Statistical Analyses $AvAl_2O_3$ and $RxSiO_2$. Exclusion 40%

	Project A				Project B				Project C			
	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2
Mt	36.46	36.84	6.44	6.14	33.63	33.82	9.66	9.83	36.88	37.22	11.10	11.00
Met	39.40	38.80	5.07	4.96	36.57	36.11	5.31	5.61	48.95	48.53	4.60	4.26
ER%	1.55		2.22		1.27		5.35		0.87		7.98	
Corr	0.80		0.05		0.99		0.98		1.00		1.00	
RR	0.64		0.00		0.98		0.97		1.00		0.99	

TABLE VII

Statistical Analyses $AvAl_2O_3$ and $RxSiO_2$. Exclusion 50%

	Project A				Project B				Project C			
	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2	Av Al_2O_3	BMU Av Al_2O_3	Rx SiO_2	BMU Rx SiO_2
Mt	36.69	37.05	6.39	6.02	33.88	34.06	9.67	9.86	37.93	38.02	10.34	10.30
Met	39.70	41.45	5.05	5.58	37.00	36.16	5.31	5.58	48.90	48.48	4.40	3.99
ER%	4.22		9.50		2.32		4.84		0.87		10.28	
Corr	0.83		-0.05		0.99		0.98		1.00		1.00	
RR	0.69		0.00		0.98		0.96		1.00		0.99	

Moreover, it is possible to establish an $AvAl_2O_3$ correlation range between 0.98 and 1 in projects B and C. However, for project A, $AvAl_2O_3$ presents a minimum correlation RR value of 0.61 and a maximum of 0.74, indicating higher variance between initial $AvAl_2O_3$ and calculated by SOM. In general, the correlation decreases as it increases the deletion to the three designs.

The $RxSiO_2$ values did not show great variations with the increase of sample exclusion. However, it is lower compared to the correlation of $AvAl_2O_3$, which is more evident for project A, where the imputation showed low correlations, with a range of RR between 0 and 0.01. For Projects B and C, the correlation $RxSiO_2$ remains high with RR between 0.95 and 0.99.

It is important to mention that the percentage errors were calculated according to the median, as it is less sensitive to fluctuations in the average values of the variable and it is more representative of heterogeneous populations, as were the groups of the original and calculated variables [15].

Discussion

This research, which seeks to improve imputation reach and interpretation of geochemical data, given actual limitations on the adequate visualization of the various data sets of high dimensionality, characterized by multiple variables. Thus, this research was conducted with SOM technique, allowing the generation of decomposed vectors, analyzed to extract the relative importance of each of the components during classification. Such an approach favored an insight into the complex relationships in sets of high-dimensional data, such as the geochemical analyses. The SOM analysis thereby favors the preservation of topological relations and, at the same time, the production of a statistical model derived from the data set [13].

Statistical analysis of the project resulted in the high correlation of $AvAl_2O_3$. About the correlation values for $RxSiO_2$ between imputed and original values in project A showed inferior results in comparison with both B and C projects, as with the results obtained for the $AvAl_2O_3$. Those results are associated to a most samples completely unrelated in regions of origin. This probably caused higher variance and uncertainty in respect of the samples of a specific variable.

Furthermore, the Project A, the maximum $RxSiO_2$ content of the original data is not a frequent value (not representative), and sometimes shows much higher levels than median. For this reason, it may be considered of analytical errors in certain samples. However, the maximum grades calculated by the SOM analysis were not directly influenced by these values. Namely, the levels calculated by SOM followed the pattern of other data $RxSiO_2$ variable.

In relation to the design B, this delivered the best results for the imputation of the data, when compared to designs A and C. These results reflect the product of the influence of a greater number of variables, regardless the diversity of origin of analyzed bauxite samples.

As for the C project, it became clear the high correlation for both $AvAl_2O_3$ and for $RxSiO_2$. However, the study was done with fewer data. In addition to this, the high value of correlation may incur in low significance, given the limited number of samples analyzed in C.

Overall, the results showed a high correlation between the values of variables measured in the laboratory and those imputed by SOM. However, in bauxite samples originating from multiple sources it was remarkable that data imputation $AvAl_2O_3$ had higher correlation with the original results to the imputation obtained for $RxSiO_2$. This can be explained by the influence of other parameters or by the absence of related variables or dependent on each other, which were not present in the project analysis involving bauxite from different regions.

Conclusion

The imputation of 1,414 pairs of values for $AvAl_2O_3$ and $RxSiO_2$ using SOM, revealed the technique as a complementary tool to generate the analytical data. Thus, SOM may be used as a tool for impute chemical composition of analytical data, besides the recognized ability of classification, integration and interpretation of multivariate data. The high correlation between the original values measured by chemical analysis in the laboratory and those imputed by SOM allowed to define the SOM effective for imputation data with up to 50% absence of values in up to two simultaneous variables.

Regarding the influence of the parameters and variables used in this study, SOM demonstrated to be more efficient when used in samples originating from nearby sources. In this case, the analyses provide the most appropriate SOM imputation, resulting in lower sampling errors. As higher the number of analytical variables input, smaller the error associated with SOM imputation data. Consequently, the analysis of other elements by instrumental analysis performance could generate greater amounts of variables, which may provide better results to the imputation, specifically in the case of $RxSiO_2$.

This study focused in two specific variables of interest ($AvAl_2O_3$ and $RxSiO_2$). Future studies may be developed to explore the imputation at higher percentages exclude values (above 50%), as well as larger amounts of variables.

Acknowledgment

The research was supported by Brazilian agency CNPq (Process 4612872014-5), in a collaboration project among Universidade de São Paulo, Universidade Federal do Pará (Brazil), Universidad Simón Bolivar (Venezuela) and Research Institute CSIRO (Australia). For imputation analyzes, this research used SiroSOM software developed by the CSIRO. We also acknowledge the Mineração Paragominas SA Company (Norsk Hydro) for laboratory and personnel support.

References

[1] S. J. FRASER, AND B. L. DICKSON, A New Method for Data Integration and Integrated Data Interpretation: Self-Organising Maps. Proceedings of Exploration 07: Fifth Decennial International Conference of Mineral Exploration, 2007.

[2] T. KOHONEN. Self-Organizing Maps. Third ed., Helsinki University of Technology Neural Networks Research Centre, Finland., 2001.

[3] T. SAMAD, S. HARP, Self-organization with partial data. Network 3, p. 205-212, 1992.

[4] I. FOLGUERA, J. ZUPAN, D. CICERONE, J. MAGALLANES. Self-organizing maps for imputation of missing data in incomplete data matrices. Chemometrics and Intelligent Laboratory Systems 143, p. 146-151, 2015.

[5] J. L. ANTONIASSI. A difração de raios X com o método de Rietveld aplicada a bauxitas de Porto Trombetas, PA. Escola Politécnica Universidade de São Paulo, São Paulo, 2010.

[6] L. SEVILLANO. Quantificação de minerais de Bauxita por difração de Raios X e sua correlação com análise química. Escola Politécnica, Universidade de São Paulo, São Paulo, 2010.

[7] C. I. E. N. Código Geológico de Venezuela. 1997.

[8] P. FREYSSINET, C. R. M. BUTT, R. C. MORRIS, P. PIANTONE. Ore-Forming Processes Related to Lateritic Weathering. Economic Geology, v. 100th Anniversary, 681, 2005.

[9] M. AUTHIER-MARTIN, G. FORTÉ, S. OSTAP, J. SEE, The mineralogy of bauxite for producing smelter-grade alumina. JOM, Mineralogy, 36, 2001.

[10] M. L. COSTA, Lateritization as a major process of ore deposit formation in the Amazon region. Exploration and Mining Geology, 6, 79, 1997.

[11] E. L. BRAY, Bauxite and Alumina. In: Mineral Commodity Summaries, USGS, 26–27, 2015.

[12] C. C. CARNEIRO, S. J. FRASER, A. P. CRÓSTA et al. Semiautomated geologic mapping using self-organizing maps and airborne geophysics in the Brazilian Amazon. Geophysics, vol. 77, no. 4, pp. K17-K24, July 1, 2012.

[13] H. S. MALEK M.A., S. S.M., AND M. I. Imputation of time series data via Kohonen self-organizing maps in the presence of missing data. Introduction to Geophysical Prospecting. 4th. Ed. McGraw-Hill. no. Engineering and Technology 41:501–506. 2008.

[14] J. VESANTO, J. HIMBERG, E. ALHONIEMI et al. SOM Toolbox for Matlab 5., H. U. o. T. Neural Networks Research Centre, Helsinki, Finland, 2000.

[15] J. C. DAVIS. Statistics and Data Analysis in Geology. John Wiley Sons, Inc., 1990.

[16] B. S. PENN. Using self-organizing maps to visualize high-dimensional data. Comput. Geosci., vol. 31, no. 5, 2005.

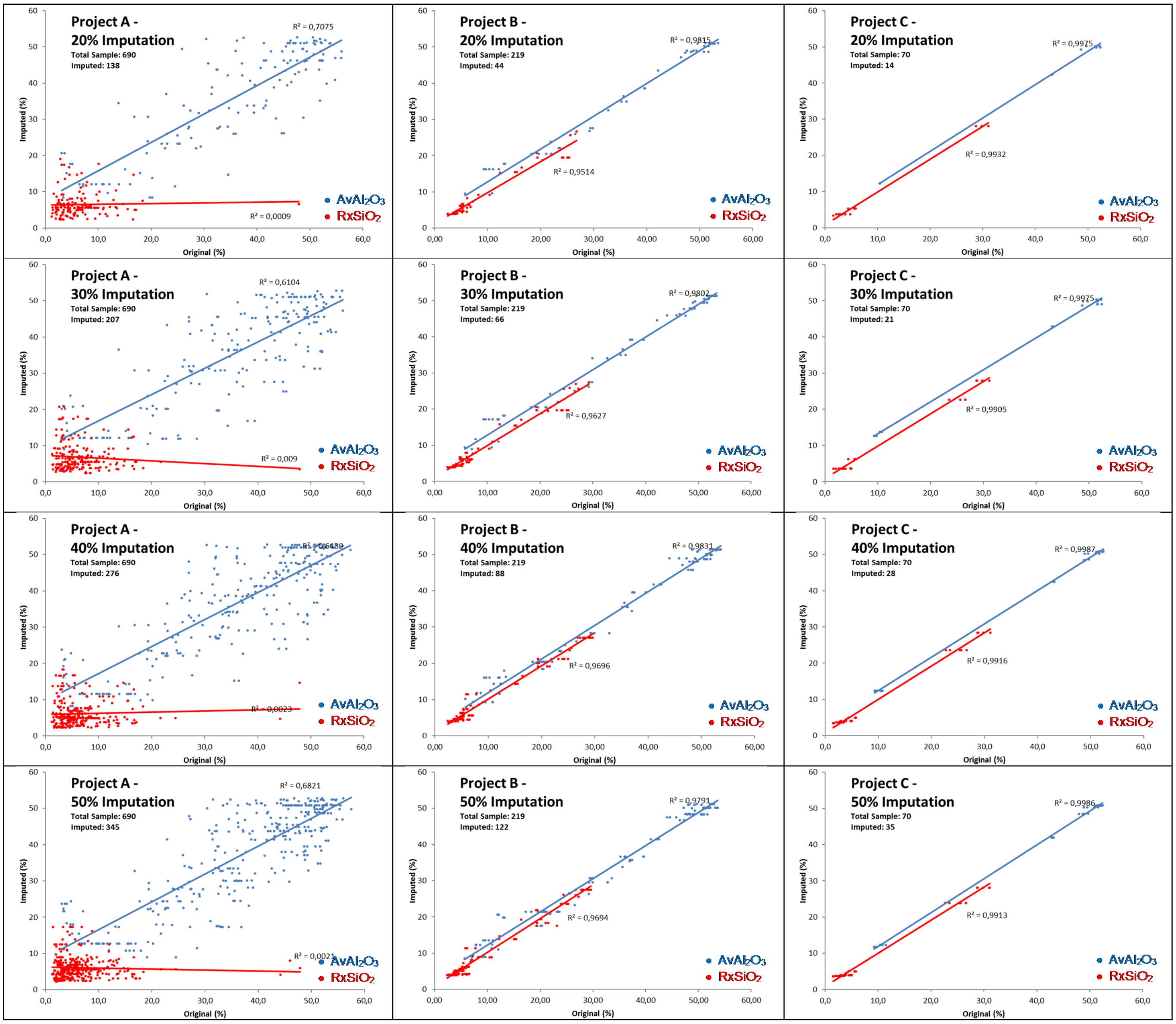

Fig. 1 Correlation between the original and imputed data in the three projects analyzed, in different amounts of sample exclusion.