Applying chemometrics to predict metallurgical niobium recovery in weathered ore

Jose Marques Braga Junior

Mestrando

Universidade Federal do Rio Grande do Sul – UFRS

Departamento de Engenharia de Minas

Porto Alegre - Rio Grande do Sul – Brasil

jose.braga@ufrgs.br

Engenheiro-Geólogo da Companhia Brasileira de

Metalurgia e Mineração - CBMM - Homogeneização

Araxá - Minas Gerais – Brasil

jose.marques@cbmm.com

João Felipe Coimbra Leite Costa

Professor Titular

Universidade Federal do Rio Grande do Sul – UFRS

Departamento de Engenharia de Minas - PPGEM-DEMIN

Porto Alegre - Rio Grande do Sul – Brasil

jfelipe@ufrgs.br

Abstract

Niobium metallurgical recovery measures how much of the metal content in the ore is separated in the concentrate after the mineral processing stages. This information can be obtained through laboratory tests with ore samples obtained during drilling. Thereunto, representative ore samples are subjected to tests mimicking the ore concentration processing flow, but these experiments are time consuming and costly.

The main objective of this study was to develop a more efficient way to obtain the metallurgical recovery information from ore samples. Based on the development of chemometrical studies, the chemical components currently analyzed in the ore with correlation to the metallurgical recovery were identified. These correlated variables were used to build a nonlinear multivariate regression model to explain the response variable, i.e. metallurgical recovery.

The Principal Component Analysis was used in this work to define which chemical variables contribute most to explain the metallurgical recovery phenomenon. The Second order regression equation (Response Surface) was the most suitable methodology to explain the metallurgical niobium recovery and was created by the interaction of the five most important chemical variables.

After the exclusion of outliers, the linear regression coefficient between the metallurgical recovery calculated and the metallurgical recovery analyzed was 82.59%.

The use of the second order regression equation contributes to reduce the amount of experimental analysis to assess the geometallurgical niobium ore response, promoting the reduction of costs for metallurgical characterization of the ore samples. The methodology proposed proved to be efficient, maintaining an adequate precision in the forecasted response.

Keywords: chemometrics, mining industry; Response Surface; geometallurgy; multivariate statistics.

1. Introduction

During the development of a mining project, in which there is mineral or metal processing to concentrate the metal of interest, it is important to know the geometallurgical response from the ore at the plant, as this is the main factor affecting mineral chain profitability. This information helps to maximize the economic benefit of the project through the integration of geological knowledge, mining, mineral processing, environmental controls and the market.

The metallurgical recovery information can be obtained experimentally at a pilot plant or by lab tests. Thus, representative ore samples are subjected to tests mimicking the ore concentration processing flow. These experiments demand time and are costly. One way to reduce the cost of these experiments is to identify chemical components currently analyzed in the ore with possible correlation with the metallurgical recovery. These correlated variables can be used to build a multivariate regression model (linear or nonlinear) to explain the response variable, i.e. metallurgical recovery.

To build and use models to predict and explain a phenomenon accurately, a careful selection is initially necessary to define the most significant variables that will explain in part, or in whole, the system's behavior. Although many variables and parameters is necessary to predict a response variable, a small number of variables, in general, explains much of it.

Thus, the first stage of a modeling process is to identify the right variables and the relationships among them.

Once the importance of modeling the geometallurgical variable is contextualized, it is proposed to investigate the relationships among the variables to explain the ore response at the plant given statistics and additive variables (grades for instance). Therefore, it is necessary to measure the variables that are considered relevant to the understanding of the analyzed phenomenon. Many are the difficulties in converting the information obtained into knowledge, especially when it comes to the statistical evaluation of such information.

Using chemometrics to calculate

the metallurgical recovery can contribute to reduce the quantity of experimental analyses of this variable, which will directly influence to reduce costs with the metallurgical characterization of the ore samples.

In this article, the development of a regression model for calculating the metallurgical niobium recovery at various stages of the production process is proposed. The results of the application are validated against the laboratory test data.

The chemical variables used in the response model are herein referred to as X_1, X_2, X_3, …, X_9 for the sake of confidentiality.

2. Material and methods

The data set considered in this study came from a drill hole campaign known as FSA. The samples obtained in this campaign were tested in the lab mimicking the processing flowchart and these samples were also chemically analyzed.

Cores from the drill holes were logged and the important geological domains were separated into: Soil; Orange weathered ore; Brown weathered ore; Saphrolite and fresh Rock (carbonatite).

Among the five geological domains mentioned, only one has enough data to be subjected to a multivariate analysis and to provide a meaningful response. This domain is the Orange weathered ore, which is the ore mined at the site.

Principal Component Analysis (PCA) helped in defining which chemical components should be considered in the multivariate regression. This math-ematical procedure uses an orthogonal transformation to convert a set of observations possibly correlated to a set of linear uncorrelated variables known as principal components. This transformation is defined so that the first principal component has the greatest variance, and each subsequent component has the maximum variance under the constraint of being orthogonal to the previous components (Pearson, 1901).

The first stage was the selection of which variables better explain the variance of the response (met recovery). Although many variables and parameters are necessary to predict the niobium metallurgical recovery, PCA analysis showed that five chemical variables (X_1, X_2, X_3, X_4 and X_5) can explain much of it. So, the other four variables (X_6, X_7, X_8, and X_9) were disregarded, as they did not have much to contribute to explain the plant metallurgical recovery. The second order multivariate regression methodology (Response Surface) was used to calculate the metallurgical recovery based on five chemical variables. This statistical methodology is commonly used for modeling and analyzing problems in which the dependent variable is influenced by several factors, and the goal is to optimize this response (George, 2007).

The Response Surface may be defined as the geometric representation obtained when a response variable is plotted as a function of two or more quantitative factors. This function can be defined as: $Y=f(x_1, x_2, …, x_k)+\varepsilon$, where Y is the response (dependent variable); and $x_1, x_2, …, x_k$ are the factors (independent variables); and ε is the random error (George, 2007).

In its polynomial equation form, the Response Surface is defined as:

$$Y=\beta_0+ \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{1=1}^{k-1} \sum_{j=2}^{k} \beta_{ij} x_i x_j + \varepsilon$$

where $x_1, x_2, …, x_k$ are the independent variables which influence the response Y; β_0, β_i (i = 1, 2, …,k), β_{ij} (j = 1, 2, …,k) are the unknown parameters; and ε is the random error. (Box and Draper, 1987).

Among the advantages of Response Surface methodology, the main one is that their results are very consistent with the effects of non-ideal conditions, such as random errors and influential points, since this methodology is robust. Another advantage is the simplicity of obtaining the polynomials from the analytical Response Surface methodology. In general, polynomials of two or more variables are continuous functions.

For multivariate regressions and the principal component analysis MINITAB® software was used, which has a number of statistical tools for both univariate analysis and for multivariate analysis. The software SGeMS was also used to build the maps and the scatter plots (Remy *et al.*, 2009).

3. Results

Among the nine variables considered for the principal component analysis (X_1; X_2; X_3; X_4; X_5; X_6; X_7; X_8 and X_9), the ones that contributed most to explain the metallurgical recovery were X_1; X_2; X_3; X_4 and X_5, which basically reduced four out of the nine initial number of independent variables. In addition to being able to identify which variables contribute most to explain a given phenomenon using PCA, it is also possible to identify which variables are redundant, which means a very high linear correlation from one variable to another.

The criterion used to make the statement that variables: X_1, X_2, X_3, X_4 and X_5 were the most contributed in this study were based on the Loading Plot analyses, which revealed the relationships among the variables considered in the space of the first two components. Figure 1 shows the plot of the loadings of the variables on the components. Each variable is a point, whose coordinates are given by the loadings on the two principal components. In this case the X_6, X_7 and X_8 variables have similar low loadings for the first component, which is different from the X_9 high loading.

The correlation between a component and a variable in PCA framework is called loading and it measures the information they share, so factor loadings represent how much a factor explains a variable in factor analysis. It is possible to examine the loading pattern in the Minitab software (Minitab, 2014) factor loading analysis output to determine the factor that has the largest effect on each variable. Some variables might have high loadings on multiple factors. Factor loadings can range from -1 to 1. Loadings close to -1 or 1 indicate that the factor strongly affects the variable. Loadings close to zero indicate that the factor has a weak effect on a specific variable.

The first component is the most important one. The variables (X_1 and X_4) provide more information, and these two variables are also strongly correlated. The variables X_6, X_7, and X_8 are closely associated and they only provide a small amount of information, as they are close to zero in the first component. Variable X_9 also has a low score in the

first component, and differently from the variables X_6, X_7, and X_8, it has high score in the second component.

According to Abdi and Williams (2010), in general, different meanings of 'loadings' lead to equivalent interpreta-tions of the components. This happens because the different types of loadings dif-fer mostly by their type of normalization.
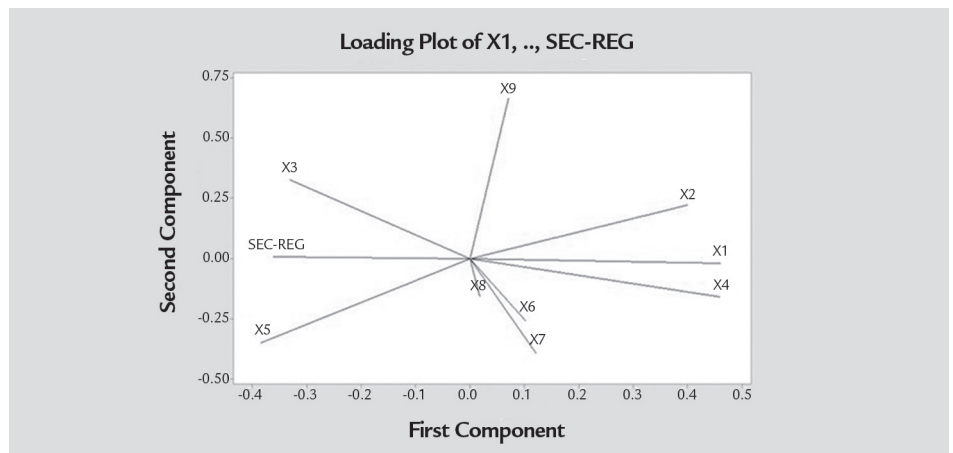

Figure 1
Factorial load chart considering nine che-mical variables and the metallurgical reco-very analyzed in the laboratory. The first load components are on the x-axis and the second load components on the y-axis.

The squared loadings could be also used to interpret the relationship between the variables. The sum of the squared coefficients of correlation be-tween a variable and all the components is equal to 1. This is due to the fact that the squared loadings give the propor-tion of the variance of the variables explained by the components.

The first multiple regression was performed without excluding outlier values. This regression was made in the MINITAB® software, and it was neces-sary to define the dependent variable (DCCG) and the independent variables (X_1; X_2; X_3; X_4 and X_5) to be used.

Figure 2 shows the result of the first regression performed without the outlier analysis. Note the scatter plots between the dependent variable and independent ones.
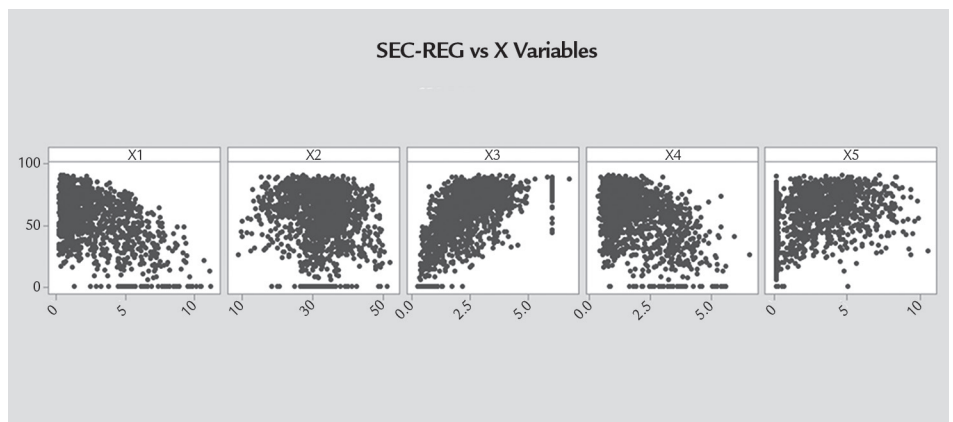

Figure 2
Scatter plots of the first attempt for multi-variable regression. These charts were built considering the independent variables (X_1; X_2; X_3; X_4 and X_5) and the result of the regression (SEC-REG). The Y-axis represents the values of the calcula-ted metallurgical recovery (SEC-REG) and the X-axis represents the value of the inde-pendent variables used in the regression

The result of the first regression without residual treatment presented a 65.04% coefficient of determination (R^2). Treating the residual appropriately can lead to a better result, optimizing the dependent variable response without in-fluencing the equation representativeness in the geological domain considered.

The residue analysis was first per-formed excluding the samples that did not contain chemical analysis for any independent variable used for regression (X_1; X_2; X_3; X_4 and X_5), i.e. keeping only an isotopic subset from the data. Missing data in the chemical analysis of all the variables needed leads to an incorrect re-sponse model. Then some specific values that were anomalous to the geological domain considered (Orange weathered ore) were excluded. Some cores were verified to understand the reasons of the geological anomaly, in most cases, the anomalies are related to veins composed of different minerals cutting the orange weathered ore. In a last stage, a residual analysis was carried out to improve the fit (R^2) of the proposed data model. No test was performed to characterize de outliers.

During the residual analysis, 202 samples were excluded, corresponding to 13% of the whole data set, where 1406 samples remained in the data set, which was used to build the regression equation.

Using the graph technique (residuals versus fitted values), the residuals that had standard deviation exceeding 35% were excluded. After the previous procedure, the linear regression coefficient between the calculated metallurgical recovery and metallurgical recovery analyzed was 82.59%. This correlation between these variables is very high, allowing the use of the regression equation to calculate the metallurgical recovery using the regression with an acceptable error.

The regression equation used to cal-culate the metallurgical recovery, after the residual treatment was defined as:

$$\textbf{SEC-REG} = -33,1 - 15,1X_1 - 0,09X_2 + 39,08X_3 + 41,91X_4 + 13,53X_5 - 2,51X_3^2 - 5,66X_4^2 - 0,62X_5^2 + 2,57X_1X_3$$

$$+ 1,49X_1X_4 - 0,23X_2X_3 + 0,23X_2X_4 - 5,16X_3X_4 - 1,63X_3X_5 - 2,78X_4X_5$$

Figure 3 shows the scatter plot between the analyzed variable (DCCG), obtained from lab tests, and the resultant metallurgical recovery using the multivariate regression equation (SEC-REG).
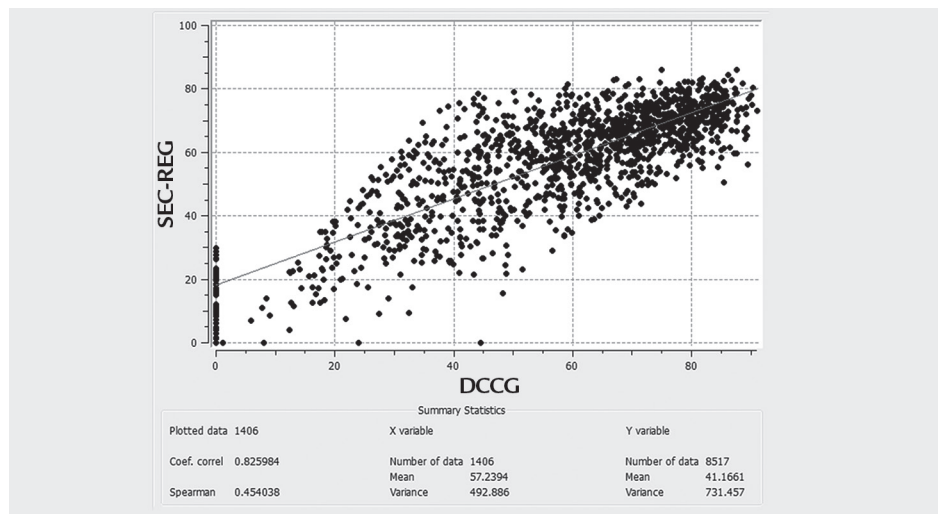

Figure 3
Scatter plot between the analyzed variable (DCCG) and the result of the calculation of the metallurgical recovery using the multivariate regression equation (SEC-REG).

Figure 4 presents the results of the multiple regression obtained with MINITAB® software. It is possible to see the sequential regression model construction with the contribution of each variable, as well as their paired variable contribution. The variable that contributes most to the model's explanation is X_3, followed by the variable X_1, and so on.
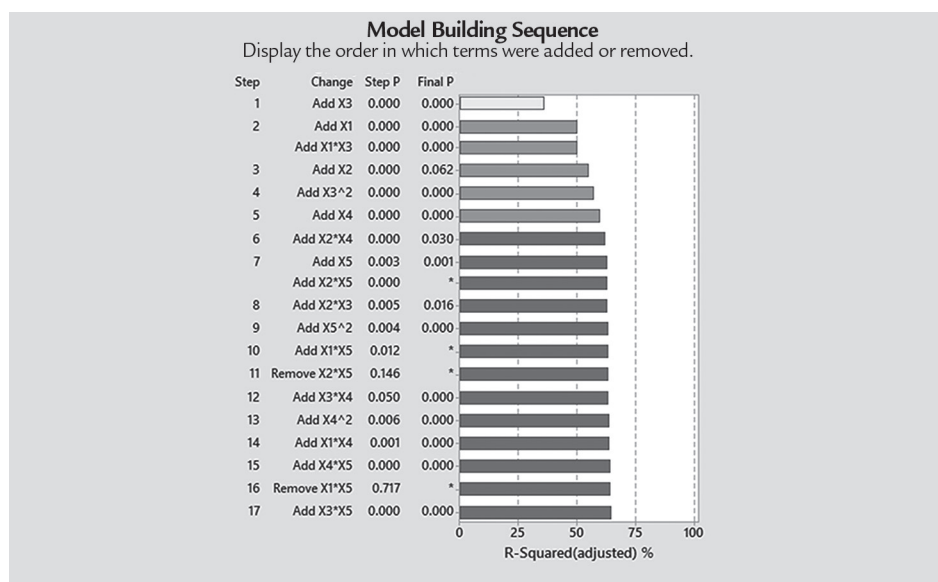

Figure 4
Summary of the model building sequence. The sequence is displayed per the contribution level of each variable or interaction of the variable pairs.

4. Discussion

The application of Response Surface methodology has been used in the chemical industry, having its foundations formalized by Box and Draper (1987). In the agronomic field, it has been used to study the yield, the effect of nutrient levels applied to the soil, including other factors such as planting density and irrigation. Although this method has been very rarely applied in mining, its use has a potential benefit. This methodology is very powerful and has many applications that can help to easily explain many phenomena involved in the mining industry.

Montgomery (2001) wrote that the Response Surface equations can be graphically represented and used in three ways as follows:

i. Describing how the test variables affect the answers;

ii. Determining the interrelationships between the variables under test; and to

iii. Describe the combined effect of all test variables on the response.

For this specific work, the target was the use of the combined effects of five variables (X_1, X_2, X_3, X_4 and X_5) to explain the metallurgical recovery of niobium.

For the effective use of Response Surface, five assumptions should be considered (Montgomery, 2001):

i. The factors that are critical to the process must be known;

ii. The region where the factors influence the process must be known;

iii. The factors must vary continuously along the chosen experimental group;

iv. There is a mathematical function that relates to the factors measured response;

v. The answer that is set by this function is a smooth surface.

Before building the regression model by the Response Surface methodology, it was necessary to perform a careful multivariate statistical data exploratory analysis. The results demonstrate it was possible to apply this methodology to find

out the effects of combining the chemical variables in the mineral recovery response during the concentration process.

Per Montgomery (2001), in the absence of sufficient knowledge about the true Response Surface, the user usually tries the first order model. However, when the first order model is not enough to fit the surface, it is necessary to incorporate higher order terms to improve the model fitted.

The use of first order regression models, with the same data set used in

this study, showed that the obtained model was insufficient to explain the behavior of metallurgical recovery from the same five chemical variables. Using this first order regression model, the best correlation obtained was 52% (R^2), which has already been improved after outlier exclusion using residual analysis.
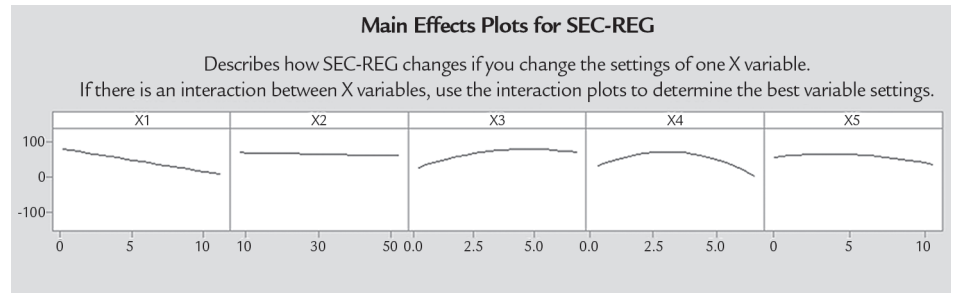
Among all the interactions between the pairs of variables, the interaction between X_4 and X_3 stands out for the phenomenon explanation, in addition to the interac-

tion between X_3 and X_1, and X_5 and X_4 that also exhibit high covariance coefficients.

The quadratic contribution of each variable also helps to define the regression model. The quadratic contribution of the X_4 is more significant in this case than the quadratic contribution of X_3. The quadratic contribution of other variables are very low.

Figure 5 shows the effect of each independent variable to the response (SEC-REG).

Figure 5
Correlation between the independent variables (X_1; X_2; X_3; X_4 and X_5) and the result of the regression (SEC-REG). Where the Y-axis represents the values of the calculated metallurgical recovery (SEC-REG) and the X-axis represents the value of the independent variables used in the regression

Main Effects Plots for SEC-REG

Describes how SEC-REG changes if you change the settings of one X variable.
If there is an interaction between X variables, use the interaction plots to determine the best variable settings.

Without the residual analysis, the regression coefficient between the analyzed and calculated metallurgical recovery was 65.04%, which means that the variables considered would explain only 65.04% of the

metallurgical recovery. Even after excluding some outliers that were affecting the regression, there were still 61 high value residues; however, they were kept to avoid creation a false impression of high correlation exclud-

ing data without well-founded criteria.

Figure 6 shows the data points with large residual (red dots in the plot) and atypical values of the response variable SEC-REG (blue dots).

Residuals vs Fitted Values

Figure 6
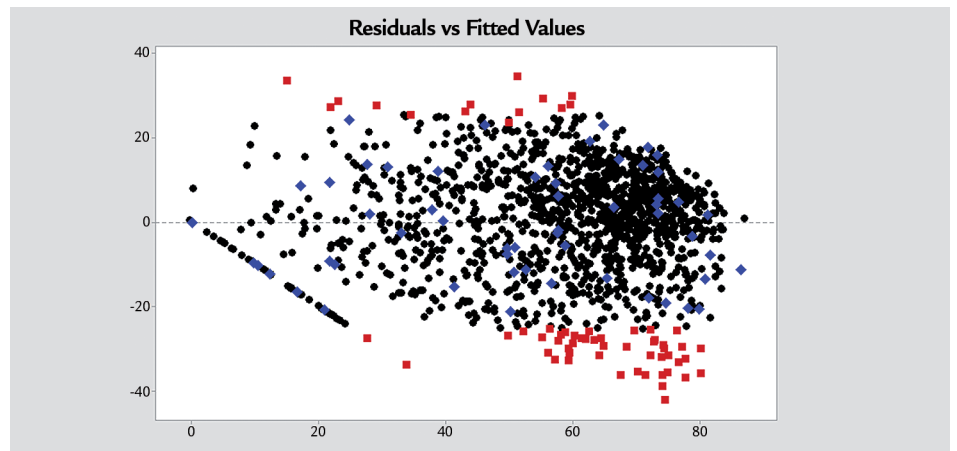Scatter plot between the residuals against the adjusted values.

Figure 7 shows two maps with the drill hole distribution: in the first map, the colored dots represent the drill holes

with geometallurgical data from laboratory tests, and the second one, with the calculated geometallurgical data. These

two data can be used together to a build simulation model in order to improve the metallurgical recovery prediction.

Figure 7
Maps showing the drill hole distribution according to the two different methodologies to obtain the processing results.

Metallurgical recovery anlized

Metallurgical recovery calculated
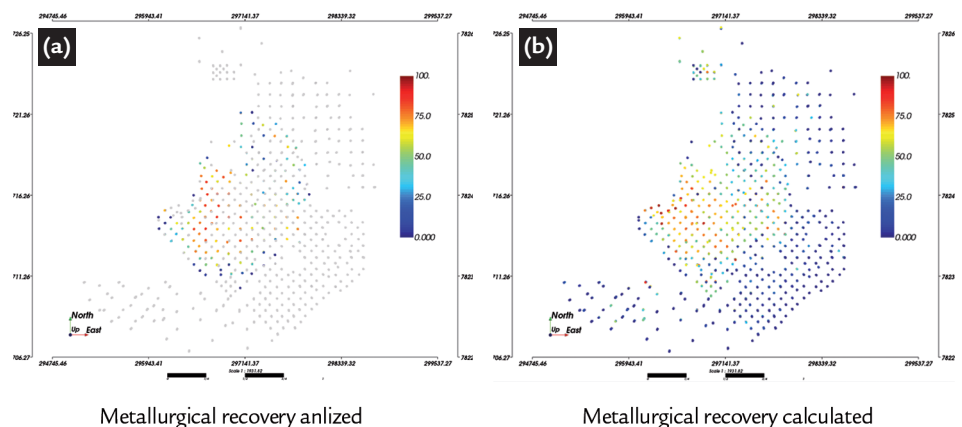
5. Conclusion

In order to create a model that explains the behavior of the metallurgical recovery, different mathematical regressions were tested. Second order multivariate regression (Response Surface) proved to be suitable for use in modeling and performance analysis of geometallurgical niobium ore, since the last is influenced by several factors.

The application of the obtained regression equation was restricted to samples associated with the geological domain "Orange weathered ore".

Before using the Response Surface methodology, it is important to define which independent variables contribute to explaining the phenomenon of interest, in this case, the metallurgical recovery of niobium.

The chemical interaction between the independent variables considered, as well as their quadratic interactions, contributed greatly to explain the phenomenon of niobium metallurgical recovery. Using the multivariate regression equation obtained, it was possible to calculate the metallurgical recovery for all samples or locations where the chemical variables are known.

The Response Surface methodology to predict the metallurgical niobium recovery provided processing information, having high correlation with the metallurgical recovery dependent data. Using this newly calculated data correctly, it is possible to increase the geometallurgical knowledge and incorporate this information into mine planning.

Acknowledgment

References

ABDI, H., WILLIAMS, L. J. Principal component analysis. **WIREs Computational statistics,** New Jersey, v.2, p. 433-459, jun. 2010.

BOX, G. E. P., DRAPER, N. R. **Empirical model building and response surface.** New York: John Wiley & Sons, 1987. 688p.

GEORGE, A. L. BOX. **Response surface, mixtures, and ridge analyses.** (2. Ed.). New Jersey: John Wiley and Sons, 2007. 1-3p.

MINITAB 17. **Getting started with Minitab: 17.** Minitab, Inc. United States, 2014. 83 p.

MONTGOMERY, D.C. **Introduction to statistical quality control.** (4. Ed.). New York: John Wiley and Sons, 2001. 816p.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine,** v. 2, n. 6, p. 559–572, 1901.

REMY, N. at alli. **Applied geoestatistics with SGeMS: A user's guide.** Cambridge: Cambridge University Press, 2009. 264p.