

# Predicting ore content throughout a machine learning procedure – An Sn-W enrichment case study

C. Iglesias<sup>a</sup>, I.M.H.R. Antunes<sup>b</sup>, M.T.D. Albuquerque<sup>c</sup>, J. Martínez<sup>d,\*</sup>, J. Taboada<sup>a</sup>

<sup>a</sup> Department of Natural Resources and Environmental Eng., University of Vigo, Spain

<sup>b</sup> ICT/University of Minho, Braga, Portugal

<sup>c</sup> Instituto Politécnico de Castelo Branco, QRural/IPCB, Portugal

<sup>d</sup> Escuela Superior de Ingeniería y Tecnología, Universidad Internacional de la Rioja, Madrid, Spain

## ARTICLE INFO

### Keywords:

Ore potential  
Machine Learning  
Classification model  
Sn-W prediction  
Stream sediments  
Portugal

## ABSTRACT

The distribution patterns of trace elements are very useful for predicting mineral deposits occurrence. Machine learning techniques were used for the computation of adequate models in trace elements' prediction.

The main subject of this research is the definition of an adequate model to predict the amounts of Sn and W in the abandoned mine area of Lardosa (Central Portugal). Stream sediment samples (333) were collected within the study area and their geochemical composition - As, B, Be, Cd, Co, Cr, Cu, Fe, Ni, P, Sn, U, V, W, Y, and Zn - used as input attributes. Different machine learning techniques were tested: Decision Trees (CART), Multilayer Perceptron (MLP) and Support Vector Machines (SVM).

For regression and clustering, CART, MLP approaches were tested and for the classification, problem SVM was used. These algorithms used six different inputs – N1 to N6 – aiming to pick out the best-performing model.

The results show that CART is the optimized predictor for Sn and W. Concerning the regression approach, correlation coefficients of 0.67 for Sn (with Input N1) and 0.70 for W (with Input N3) were obtained. Regarding the classification problem, an error rate of 0.10 was reached for both Sn (Input N1) and W (Input N2).

The classification process is the best methodology to predict Sn and W, using as input the trace element concentrations in the collected stream sediment samples, Lardosa area, Portugal.

## 1. Introduction

Distribution models are generally used to predict the presence or absence of target deposits in a specific area that has not been surveyed or censused (Bahn and McGill, 2013).

Stream sediments were collected and represent a non-random point distribution. The occurrence of a set of deposits, represented as points, can be characterized as dependent variables, which can be explained by geological, geographical, and geospatial data (Boots and Getis, 1988; Diggle, 1983, 1990; Rowlingston and Diggle, 1991).

The mineralogical content of stream sediments from mineralized areas is primarily linked to the gangue constituent's mineralization or minerals from host rocks. The wind as the predominant process may be responsible for the verified anomalous high concentrations (Santos Oliveira et al., 1998). The stream sediments correspond to heterogeneous materials resulting from the superficial weathering and transported by river flow from the source rock to a point of deposit (Porwal et al., 2003).

The spatial distribution patterns of trace elements can aid in the prediction of ore deposits and mineral occurrences. There has been much research on modeling distribution patterns using environmental and geospatial data to make predictions based on the projection of predictive attributes of geologic and geographic data from models such as global circulation models (Iverson and Prasad, 1998; Strebelle, 2002; Skov and Svenning, 2004).

A mineral distribution model is important for identifying potential mining areas for economic development and soil management (Scott and Csuti, 1997; Ferrier, 2002). Finding new mineral deposits involves the formation, transportation, and deposition of mineralized fluids (Falconer, 1912; Bowden and Jones, 1978).

Spatial autocorrelation is a phenomenon where the presence of some property in a sampling unit makes its presence in neighboring sampling units likely (Harris et al., 2010; Ibrahim and Bennett, 2014).

Different researchers have used machine learning (ML) as a tool for modeling geospatial data by building a model capable of identifying patterns in geospatial data and making predictions from these patterns

\* Corresponding author.

E-mail address: [javmartinez@uvigo.es](mailto:javmartinez@uvigo.es) (J. Martínez).

e.g. (Porwal et al., 2003; Oommen et al., 2008; Leite and de Souza Filho, 2009; Oh and Lee, 2010; Carranza, 2011; Ballabio and Sterlacchini, 2012; Ibrahim and Bennett, 2014; Daszykowski et al., 2015; Rodriguez-Galiano et al., 2015; Chen et al., 2014; Chen and Wu, 2017; Zuo and Xiong, 2018; Zuo, 2017; Zhao et al., 2016). Also, different studies are in the bibliography related to ore: formation of ore (Duan et al., 2018) and ore genesis (Zhang et al., 2018; Fan et al., 2016) but not to predict its content.

The aim of this research work is to construct a model to predict the amount of Sn and W in stream sediments associated with abandoned mining activities at Lardosa (in Central Portugal).

Different machine learning techniques were tested: decision trees (CART), multilayer perceptron (MLP) and support vector machines (SVM) for regression and clustering, and CART, MLP, and SVM for supervised classification.

Prediction algorithms were implemented, namely using regression and classification approaches, for a comparison of results.

## 2. Materials and methods

### 2.1. Study area and dataset

The study mining area is approximately 130 km<sup>2</sup> located in the district of Castelo Branco in central Portugal (Fig. 1a). Geologically, the area comprises the Iberian Massif in the Central Iberian Zone (CIZ). The prevailing rocks are schists and greywackes from the Beiras Group (Schist-greywacke Complex), which were intruded by the Castelo Branco and Oledo granitic plutons (Antunes et al., 2009). NW-SE to WNW-ESE quartz veins with cassiterite and wolframite veins fill late to post-tectonic Variscan faults following the Sn-W mineralization from Góis-Segura. The Sn-W quartz veins also contain frequently associated Fe, Cu and As sulphides, related to the granitic intrusion. Younger quartz veins rich in galena and sphalerite cut the schist-metagraywacke complex, and a few of them intrude the granite. The Lardosa deposits also characterize the region, resulting from the alteration of Lardosa granite and consequent mobility and accumulation of cassiterite and ilmenite (Instituto Geológico e Mineiro, 1998; Antunes et al., 2014). The geological background makes this mining area interesting, as it contains 11 abandoned mine concessions containing W, Sn-W, and Zn-Pb (Antunes et al., 2014).

**Table 1**  
Analytical conditions used for trace elements determination (Instituto Geológico e Mineiro, 1998).

Analytical method	Element	Detection limit	Precision
ICP-AES	As	20 ppm	20 ppm
	B, Co, Cr, Cu, Ni, P, V, Zn	10 ppm	10 ppm
	Be, Cd	1 ppm	10 ppm
	Fe	0.10%	10 ppm
	U	0.5 ppm	10 ppm
	Y	5 ppm	10 ppm
	X-Ray fluorescence spectrometry	Sn, W	10 ppm

The geochemical composition of stream sediments comprised 333 samples (Fig. 1b), collected during the period 1980–1988 in a narrow region, ranging from 50 m upstream to 100 m downstream from the stream confluences (Instituto Geológico e Mineiro, 1998).

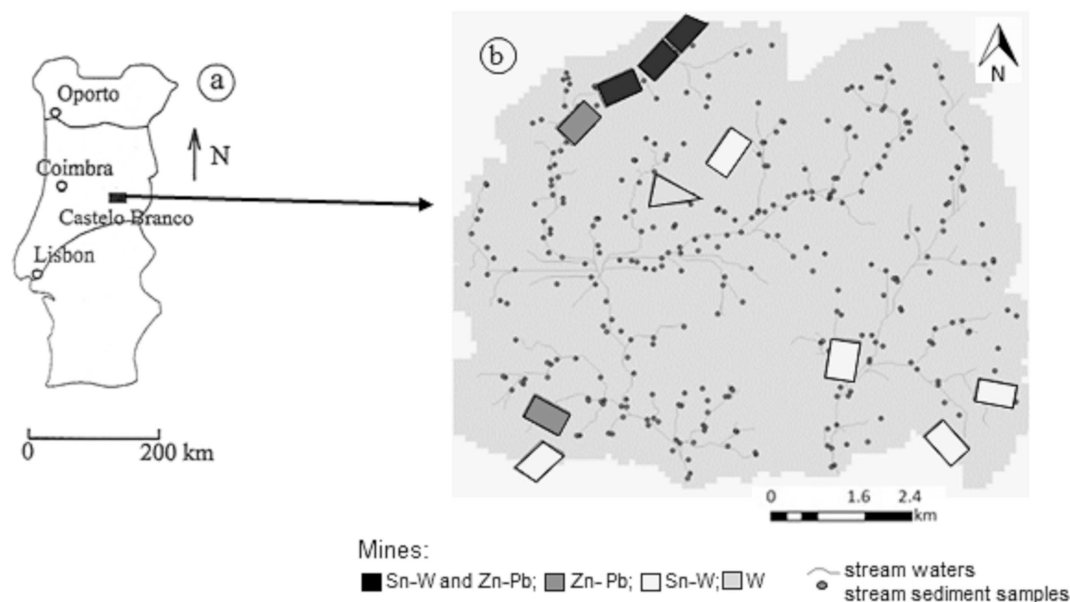
All the samples were prepared through reduction, drying and grinding. Total concentrations of As, B, Be, Cd, Co, Cr, Cu, Fe, Ni, P, V, U, Y, and Zn, in the fraction <74 µm, after digestion with Aqua-Regia were determined by ICP-AES. Tin and W were analyzed by X-ray fluorescence spectrometry (Instituto Geológico e Mineiro, 1998) and were used for the prediction model fitting.

The main laboratory conditions of the analyzed elements included in the dataset are presented in Table 1. The box plots obtained for the sixteen trace elements analyzed and details of them are presented in Figs. 2 to 4. All the box plots are displayed with the median of the data represented by the central line of the box, the upper and lower edge corresponding to 75th and 25th percentiles, respectively, and the whiskers indicating the range outside which data points are considered outliers.

The descriptive statistics obtained for the trace elements are shown in Table 2. Tin and W present a similar variability according to their average and median. However, stream sediments from the study area contain higher concentrations of Sn (up to 138 ppm) than W (Table 2).

### 2.2. Machine learning techniques used

Machine learning techniques are used in a broad variety of



**Fig. 1.** Study area: a) geographical location on the map of Portugal; b) stream sediment samples. Adapted from (Antunes et al., 2014).

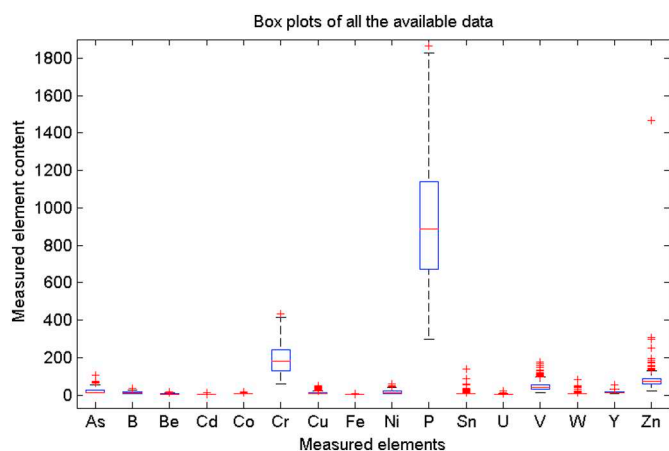


Fig. 2. Box plots of the sixteen trace elements.

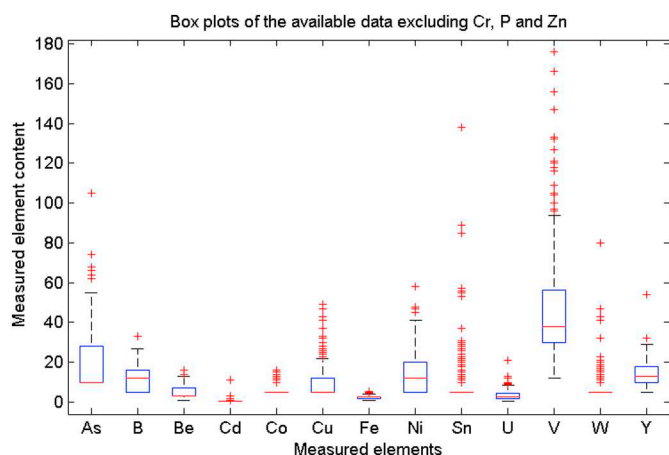


Fig. 3. Detailed box plots for trace elements, except for Cr, P and Zn.

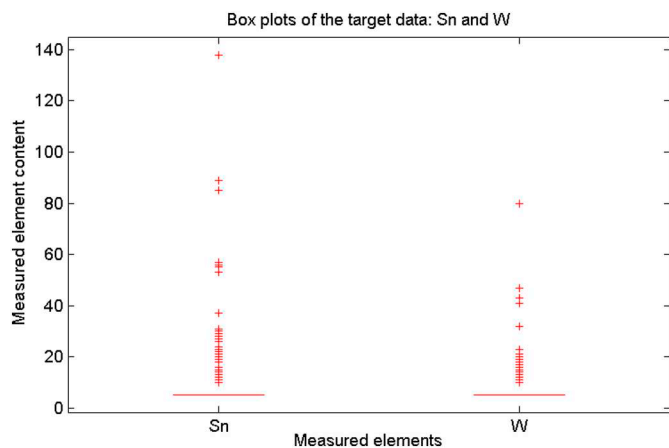


Fig. 4. Selected box plot for Sn and W.

applications (Porwal et al., 2003; Leite and de Souza Filho, 2009; Oh and Lee, 2010; Carranza, 2011; Ibrahim and Bennett, 2014; Rodriguez-Galiano et al., 2015). The main goal of this research is the prediction of Sn and W, in an abandoned mining area at Lardosa (Central Portugal) and the following machine learning techniques will be applied.

Regardless of the technique used to analyze the data, a 10-fold cross-validation was performed to optimize the results. The initial dataset was divided into ten subgroups; nine of them were used in the training stage and the remaining one for testing. Finally, the process

Table 2

Descriptive statistics for trace element concentrations (ppm).

Element	Minimum	Maximum	Average	Median	Variance	Standard deviation
As	10.0	105.0	19.7	10.0	217.2	14.7
B	5.0	33.0	11.4	12.0	28.6	5.4
Be	1.0	16.0	4.8	3.0	10.7	3.3
Cd	0.5	11.0	0.6	0.5	0.4	0.6
Co	5.0	16.0	5.5	5.0	3.4	1.8
Cr	61.0	431.0	190.4	181.0	5525.9	74.3
Cu	5.0	49.0	9.3	5.0	63.0	7.9
Fe	0.9	5.2	2.5	2.5	0.6	0.8
Ni	5.0	58.0	14.5	12.0	100.1	10.0
P	299.0	1865.0	928.5	885.0	101,359.2	318.4
Sn	5.0	138.0	8.7	5.0	146.5	12.1
U	0.5	21.0	3.5	2.9	5.1	2.3
V	12.0	176.0	46.9	38.0	733.9	27.1
W	5.0	80.0	7.2	5.0	42.0	6.5
Y	5.0	54.0	14.0	13.0	36.4	6.0
Zn	19.0	1465.0	82.4	73.0	6911.8	83.1

was repeated through 10 iterations, incorporating the contributes of all the subgroups considered.

2.2.1. Decision trees (CART)

Given a set of continuous or categorical independent variables, classification and regression trees (CART) are non-parametric techniques that can explain the response of a dependent variable (Breiman et al., 1984; Moisen, 2008). The initial data are split into subgroups, then the variance of these subgroups is minimized until they reach homogeneity (Brown and Myles, 2009).

Variance minimization can be accomplished using different algorithms such as mean-squared error (MSE); Gini's diversity index, deviance or towing rule. In this research paper, regression trees were optimized using MSE, while classification trees were optimized using Gini's diversity index. The Gini index at node *t* can be expressed as:

$$\sum_{j \neq i} p(j/t)p(i/t) \tag{1}$$

where *i* and *j* represent the categories of the target variable (dependent variable), and *p<sub>i</sub>* *t*(*j*) are the proportion of target categories *j* and *i*, respectively, present in the node *t*

$$1 - \sum_j p^2(j/t). \tag{2}$$

Gini's index is usually used for categorical dependent variables and its value ranges from 0 (maximum equality) to 1 (maximum inequality) (Soman et al., 2006).

The main advantage of using CART is the model optimized fitting, considering the original data. Moreover, it allows a quick, visual and intuitive results' interpretation through the tree observation. The presence of outliers is minimized through pruning (Breiman et al., 1984).

2.2.2. Clustering

The basis of this unsupervised classification method is the ability to differentiate the elements of a given group based on a specific characteristic, usually a distance measure (Jain et al., 1999; Kotu and Deshpande, 2015). This methodology is appropriate for data with limited information since clustering explores the interrelationships among the data points to assess their structure (Jain et al., 1999).

The methodology could be applied to classification problems, and the main steps are the following (Jain et al., 1999): (i) the initial pattern is represented (ii) define a pattern of the proximity of the data; then (iii) an initial clustering is applied and (iv) the similarity of the elements of each dataset is analyzed. This sequence is repeated until the elements of a subgroup (or cluster) have the maximum similarity while elements of different subgroups have the minimum similarity.

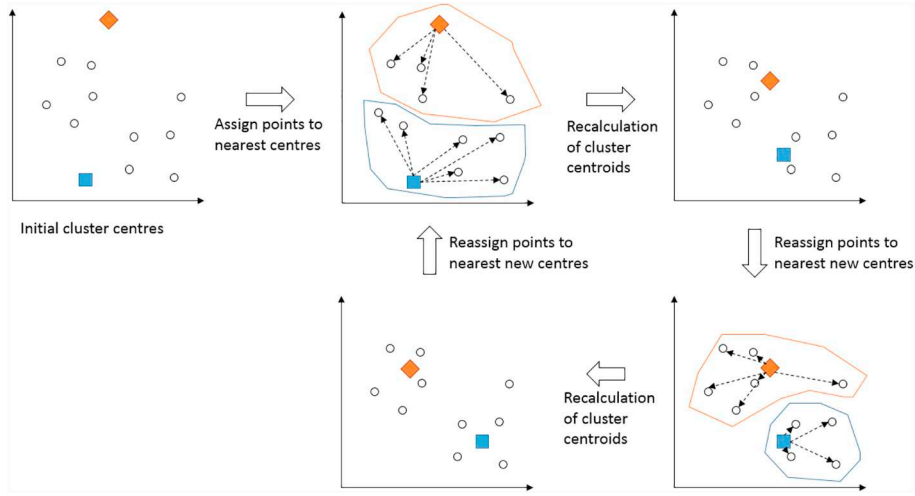


Fig. 5. Example for k-means clustering algorithm.

Several algorithms can be used for measuring similarity and all of them are based on the distance between the elements in N-dimensional space. Likewise, different approaches to clustering data can be invoked (Jain et al., 1999). In this research, the squared error criterion and k-Means clustering algorithm were used.

Considering a clustering  $L$  of a pattern set  $X$ , that contains  $K$  clusters, the squared error is expressed as

$$e^2(X, L) = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i^j - c_j\|^2, \tag{3}$$

being  $x_i^j$  the  $i$ th pattern that belongs to the  $j$ th cluster, and  $c_j$  the centroid of the  $j$ th cluster. Using a random partition, the algorithm evaluates the similarity between the pattern (or element) and the cluster center until these converge (Fig. 5).

Clustering was applied as a semi-supervised technique in this study, setting the number of clusters to be tested by the model. Thus, the algorithm determines the centroid of each cluster based on the distances between it and the items of the dataset.

### 2.2.3. Multilayer perceptron (MLP)

Neural networks (McCulloch and Pitts, 1943; Bishop, 2008) are a machine learning technique characterized by a layered structure and each layer containing a certain number of nodes (or neurons). There are three types of layers: input layer (containing the initial data), output layer (containing the required response) and hidden layers (Fig. 6).

Each layer is composed of: (i) neurons with input connections and weights that regulate the intensity of input signals, (ii) an activation threshold, (iii) an activation function that depends on the input signals and (iv) an output as a function of the input signal, also called “transfer function”. The weights that connect the neurons of the different layers are calibrated with a training algorithm based on the deviation of the

network outputs with respect to the real values.

In a  $d$ -dimensional input space with a  $c$ -dimensional output layer, the neural network implements a function that can similarly be expressed equally

$$\begin{aligned} f(x) &= \phi(\psi(x)) \\ \phi: X \subset R^d &\rightarrow T \subset R^p \\ \psi: T \subset R^p &\rightarrow Y \subset R^c \end{aligned} \tag{4}$$

where  $T$  is the hidden layer and contains  $p$  neurons; it is also known as “characteristic space”.

Multilayer Perceptron (MLP) is a sensitive case of neural networks where neurons are perceptrons and the learning process is characterized by backpropagation (Bishop, 2008; Lek and Park, 2008). The backpropagation algorithm finds the optimum combination of squared errors and weights and generates the best output from the trained network. The implemented function is (Heaton, 2012)

$$f(x) = \sum_{j=1}^p \phi_j(c_j \psi(w_j^T x + w_0) + c_0) \tag{5}$$

being  $c_j$  and  $c_0$  the weights of the hidden layer, and  $w_j, w_0$  the weights of the input layer.

### 2.2.4. Support vector machines (SVM)

The support vector machines (SVM) are a methodology initially proposed by Cortes and Vapnik (1995) and applied to a binary classification process. Since then, support vector machines have been widely used in pattern recognition due to their predictive ability, flexibility, parsimony and global optimum capacity (Borges, 1998). SVM is usually applied to classification problems, although good performances have also been obtained in regression problems and time series applications (Thissen et al., 2003), is commonly referred to as support vector

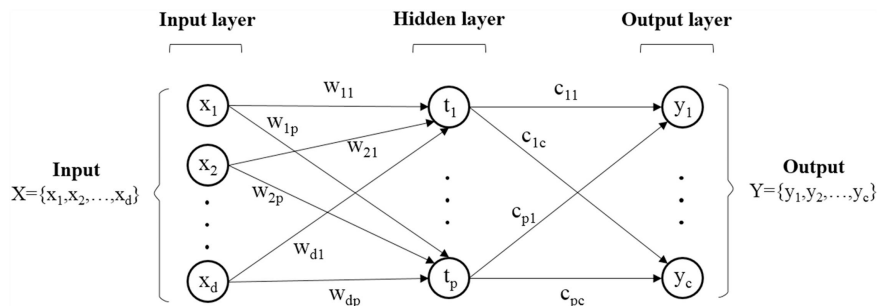


Fig. 6. Illustration of the structure of a neural network with  $d$  input neurons,  $t$  hidden neurons and  $c$  output neurons.

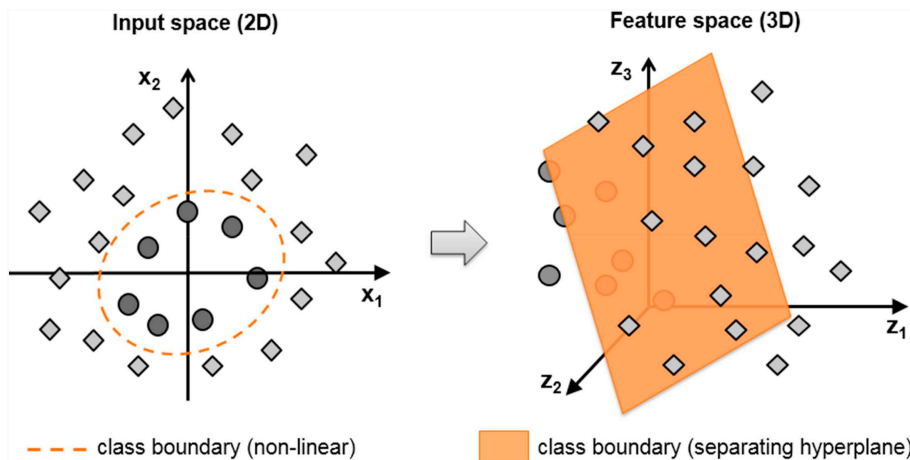


Fig. 7. Illustration of the transformation of a 2D-input space (left) into a 3D-feature space (right) by means of the kernel function.

regression (SVR) (Smola and Schölkopf, 2004).

Basically, given a training data, it being  $\chi$  a  $d$ -dimensional input space, the objective of SVR is to find a function  $f(x)$  with a maximum deviation  $\epsilon$  from the obtained targets  $y_i$  for all the training data (Cortes and Vapnik, 1995). If we considered a linear function  $f$

$$f(x) = \langle w, x \rangle + b \text{ with } w \in \chi, b \in \mathbb{R} \quad (6)$$

where  $\langle w, x \rangle$  is the dot product in  $\chi$ , the optimization problem can be written as

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 \\ & \text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon \\ \langle w, x_i \rangle + b - y_i \leq \epsilon. \end{cases} \end{aligned} \quad (7)$$

Bennett and Mangasarian (1992) proposed a soft margin loss function to allow a certain number of errors in the resulting output, thus rewriting the previous equations as

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ & \text{subject to } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (8)$$

where  $\xi_i, \xi_i^*$  are slack variables and  $C > 0$  represents the tolerated deviations, larger than  $\epsilon$ .

Considering nonlinear functions, the dualization method presented in (Fletcher, 1989) uses Lagrange multipliers to rewrite the optimization problem as the ‘‘Support Vector expansion’’

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i, \text{ thus } f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b, \quad (9)$$

where  $\alpha_i, \alpha_i^*$  are the Lagrange multipliers. With the latter equation,  $w$  is completely described as a linear combination of the training patterns  $x_i$ .

Considering a classification problem, its resolution relies on the concept of the optimal separating hyperplane introduced by Vapnik (1982). Given a set  $z^n$  where  $z_i = (x_i, y_i)$  is a sample with  $x_i \in \chi \subset \mathbb{R}^d$ ,  $y_i \in Y = \{-1, 1\}$ ,  $i = 1:n$  linearly separable by a decision function  $f_w, b(x) = \text{sign}(\langle w, x \rangle + b)$ ,  $w \in \mathbb{R}^d, b \in \mathbb{R}$ .

Then, the optimal separating hyperplane is the one whose distance to the classes is maximum (the margin is maximum). The margin of a hyperplane  $\tau_{w,b} = \{x \in \chi / \langle w, x \rangle + b = 0\}$  with respect to the sample  $z^n$  is calculated as

$$\gamma(\tau_{w,b}) = \min_{i \in \{1:n\}} \frac{1}{\|w\|} |\langle w, x \rangle + b| = \frac{1}{\|w\|} \min_{i \in \{1:n\}} |\langle w, x \rangle + b| = \frac{1}{\|w\|} \chi(\tau_{w,b}), \quad (10)$$

being  $\chi(\tau_{w,b})$  the functional margin. So, the optimal hyperplane is obtained by solving the problem

$$\begin{aligned} & \max_{w \in \mathbb{R}^d, b \in \mathbb{R}} \{\gamma(\tau_{w,b}) = \min_{i \in \{1:n\}} |\langle w, x \rangle + b|\} \\ & \text{under: } \|w\| = 1. \end{aligned} \quad (11)$$

Now, considering the normalized functional margin and the dualization of the problem, the support vectors provide the solution (Vapnik, 1998; Cristianini and Shawe-Taylor, 2000)

$$w = \sum_{i \in SV} \beta_i x_i \Rightarrow f_{w,b}(x) = \sum_{i \in SV} \beta_i \langle x_i, x \rangle + b. \quad (12)$$

The concept of soft margin is introduced to admit several misclassified observations. The linear problem is usually insufficient since the analyzed sample is frequently non-linearly separable. This issue is tackled with the kernel trick, which transforms the input space based on the following theorem: ‘‘if  $\chi$  is a topological space and  $k$  is a continuous positive definite function in  $\chi \times \chi$ , then there is a Hilbert space  $\mathcal{H}$  and a continuous application  $\phi: \chi \rightarrow \mathcal{H}$  such that  $\forall x, x'$  verifying  $k(x, x') = \langle \phi(x), \phi(x') \rangle$ ’’ (Fig. 7).

The solution of the classification problem can be written as

$$\begin{aligned} w &= \sum_{i \in SV} \beta_i \phi(x_i) \Rightarrow f_{w,b}(x) = \sum_{i \in SV} \beta_i \langle \phi(x_i), \phi(x) \rangle + b \\ &= \sum_{i \in SV} \beta_i k(x_i, x) + b. \end{aligned} \quad (13)$$

### 2.3. Methodology

The prediction of Sn and W content in the abandoned mineralized area of Lardosa (Central Portugal) was approached in two different ways. Firstly, the regression problem was considered, while a second phase was based on the result of a categorization problem.

The regression methodology considers the original Sn and W concentrations as the output variables while the classification approach needs the definition of several classes of the target variables. These categories were determined considering both the distribution of Sn and W concentrations and their geochemical behavior. The adopted criterion was the definition of different anomaly limits: the first limit representing the average distribution and the subsequent ones representing the low, medium, and high anomaly level, respectively.

Stream sediments have a large capacity to retain trace elements and their chemical composition can be a good indicator of associated mineralization. However, Sn and W have relatively low mobility and tend



**Table 3**  
Defined Sn and W intervals and relatively frequency for the classification problem.

Class	Sn (ppm)		W (ppm)	
	Interval	Frequency	Interval	Frequency
1	$x < 8$	254	$x < 9$	264
2	$8 \leq x < 31$	69	$9 \leq x < 21$	62
3	$31 \leq x < 60$	7	$21 \leq x$	7
4	$60 \leq x$	3	–	–

to be retained in mine dumps and mine tailings as opposed to being distributed through stream networks. The defined classes for Sn and W initial concentrations and the number of observations for each one is presented in Table 3.

The first dataset (Input N1), comprised all the analyzed trace elements (As, B, Be, Cd, Co, Cr, Cu, Fe, Ni, P, U, V, Y, and Zn) as input variables.

Input N2, based on previous studies carried out in the same area showed that As, Cd, Co, Cr, Cu, Fe, Ni, V and Zn stream sediments concentrations are related to the Sn-W and Pb-Zn mineralization from the area while P, B, Be, Y and U stream sediments concentrations. These results led to the selection of As, Cd, Co, Cr, Cu, Fe, Ni, V, and Zn, as input N2.

Principal Component Analysis (PCA) (Wold et al., 1987; Jackson, 1991) was undertaken as an exploratory study, aiming the definition of association patterns within the dataset and the definition of the relative strength of each element in the final computation. Input N3 corresponds to the first Principal Plan (Factorial Axes 1 and 2), explaining 95% of the total variance and are most heavily weighted by P and Zn. In a second step, a new PCA was performed focused on the nine trace elements, mostly related to the Sn-W and Pb-Zn mineralization, in the area. Chrome, V, and Zn explained > 98% of the total variance and both contributing to the first factorial component. Input N4 includes these three trace elements.

The previous research work by Antunes et al. (2014) studied the environmental risk associated with W-Sn and Pb-Zn mineralization in this area applying geostatistical methodologies to create risk maps using the Geoaccumulation Index (Igeo) of trace elements from stream sediments. PCA was applied to Igeo values and two synthesis variables were identified:

- (i) F1: this factor showed the dependence of P and B and their inverse correlation with a cluster formed by Cr, Ni, and V.
- (ii) F2: this second factor showed the dependence of As, Fe, and Zn.

With respect to these findings, the remaining input datasets were defined as follows: Input N5 included the elements derived from F1 (P, B, Cr, Ni, and V), while Input N6 comprised the elements of both F1 and F2 (P, B, Cr, Ni, V, As, Fe and Zn).

Summing up and to clarify the steps of our methodology, the different configurations of the predictive model included the elements of the six inputs summarized in Fig. 8. These inputs were used for both the regression and the classification experiments. In the case of regression, Sn and W data are the original concentrations of stream sediments, while in the case of classification, the classes included in Table 3 are the desired output of the model.

Considering the machine learning techniques, CART, MLP, and SVM were tested in the regression experiment, whereas the classification experiment also included clustering. The network architecture for MLP both for regression and classification was as follows: one input layer, one hidden layer, and one output layer; the transfer function was the Log-sigmoid and the maximum train epochs were 50,000 (although this number was never reached since the method converged before that).

The performance of the models was assessed based on the coefficient of determination  $R^2$  in the case of regression, which represents the proportion of total variation of the predicted variable explained by the model. In other words, the coefficient of determination indicates the proximity between real data points and a regression curve. Its values range from 0 to 1, being  $R^2 = 1$  if the approximation is perfect and  $R^2 = 0$  in the opposite case. More details can be found in (Freedman et al., 2007; Iglesias et al., 2014).

In the case of the classification approach, this assessment was based on the obtained error rates (proportion of misclassified elements with respect to the total number of elements of a certain subset). The resulting values range from 0 to 1; 0 indicates that all the observations are incorrectly classified whereas 1 indicates the perfect fit. Therefore, the error rate is the inverse of the accuracy. Additionally, confusion matrixes can be calculated and other evaluation parameters such as precision and recall can be used to assess the performance of the models. They are defined as follows:

$$\text{Precision} = \frac{\text{Obs. correctly classified as A}}{\text{Total obs. classified as A}}$$

$$\text{Recall} = \frac{\text{Obs. correctly classified as A}}{\text{Total real obs. from A class}} \quad (14)$$

### 3. Results and discussion

As mentioned above, the resolution of the problem was tested using both regression and classification approaches. The results obtained for each block are shown in detail below.

#### 3.1. Results of the regression problem

The regression problem includes the results for Sn and W prediction by means of the CART, MLP, and SVM from the input data. These techniques were evaluated for predicting Sn and W values individually using the 6 different input datasets described. The determination coefficients obtained for Sn and W predictions are summarized in Tables 4 and 5, respectively. The CART methodology achieves the best predictions for Sn and W.

In this study, it is important to determine the optimal parameters of each model and determine those potential overfitting problems. In the case of the MLP, the optimum number of neurons ranged from 5 to 18 in the different tests (both for the regression and classification approaches), noting that from those values overfitting problems appeared.

The best Sn prediction ( $R^2 = 0.67$ ; Table 4) was obtained with the input N1, comprising all the analyzed trace elements (Input N1 = {As, B, Be, Cd, Co, Cr, Cu, Fe, Ni, P, U, V, Y, Zn}). Tungsten prediction gives slightly better results ( $R^2 = 0.70$ ; Table 5) with Input N3 = {P, Zn}.

#### 3.2. Results of the classification problem

This section presents the results of Sn and W prediction obtained from a different approach: a classification problem. Firstly, the definition of four categories was performed for Sn, and three for W, according to the previously defined intervals (Table 3). Hence, the model output corresponds to a specific category (1, 2, 3 or 4) instead of a given value for the selected elements.

The majority Sn and W observations belong to the first class, with a low concentration of Sn (~8 ppm) and W (~9 ppm) (Table 3). Considering the distribution of classes, 4 for Sn ( $\geq 60$  ppm) and 3 for W ( $\geq 21$  ppm), only a few observations belong to these intervals (Table 3). The limited number of each class frequency influences the training of the model, particularly the possible defined number of subsets. For Sn, a maximum of 3 subsets could be defined, while a maximum of 7 subsets is possible for W (Tables 6 and 7).

The best predictions are achieved with a CART as was obtained in the regression problem. In the classification problem, error rates about

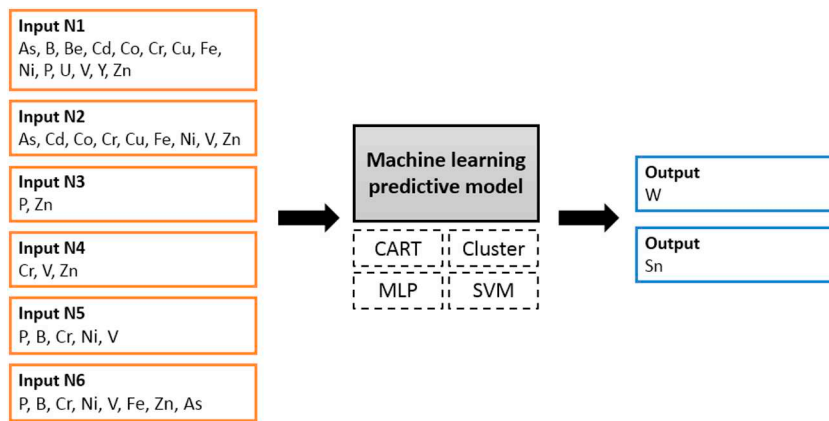


Fig. 8. Summary of the six different input sets and machine learning techniques tested.

**Table 4**  
Determination coefficients obtained for Sn prediction.

Output data	Input data	Regression problem (R <sup>2</sup> )		
		CART	CART	CART
Sn	Input N1	<b>0.67</b>	0.20	0.15
	Input N2	0.58	0.03	0.09
	Input N3	0.63	–	0.11
	Input N4	0.54	0.05	0.09
	Input N5	0.54	0.27	0.07
	Input N6	0.64	0.26	0.06

The best result is shown in bold.

**Table 5**  
Determination coefficients obtained for W prediction.

Output data	Input data	Regression problem (R <sup>2</sup> )		
		CART	MLP	SVM
W	Input N1	0.58	0.05	0.10
	Input N2	0.65	0.02	0.07
	Input N3	<b>0.70</b>	0.03	0.05
	Input N4	0.47	0.05	0.09
	Input N5	0.53	0.02	0.05
	Input N6	0.48	0.03	0.07

The best result is shown in bold.

10% for both Sn and W were obtained (Tables 6 and 7). The 6-input data tested, perform similarly for Sn and W prediction, with error rates ranging from 10% to 15% and an accuracy of 85% to 90%. As has been

**Table 6**  
Error rates for Sn predicted with Cluster, CART, MLP and SVM methodologies.

Output data	Input data	Classification problem							
		Cluster		CART	MLP		SVM		
		Training error rate	Test error rate	Error rate	Training error rate	Test error rate	Training error rate	Test error rate	
Sn	Input N1	0.72	0.85	<b>0.10</b>	0.10	0.30	0.00	0.23	
	Input N2	0.77	0.83	0.15	0.10	0.30	0.00	0.23	
	Input N3	0.24	0.24	0.13	0.20	0.25	0.23	0.23	
	Input N4	0.76	0.81	0.15	0.20	0.25	0.01	0.26	
	Input N5	0.77	0.78	0.12	0.20	0.25	0.00	0.23	
	Input N6	0.78	0.92	0.12	0.10	0.30	0.00	0.23	

The best result is indicated in bold.

obtained with the regression problem (Table 4), Sn was best predicted with Input N1, considering all the trace elements (R2 = 0.10; Table 6). For W, the best prediction (R2 = 0.09; Table 7) was obtained with Input N2 = {As, Cd, Co, Cr, Cu, Fe, Ni, V, Zn}. The results are consistent with the a priori expectation since the optimized model fitting is observed when using the N1input set, which has the largest number of attributes, and, therefore, more information.

Given the obtained results and considering that this is a classification problem, confusion matrixes were calculated for Sn and W prediction (Tables 8 and 9). Precision and recall were also estimated to complete the appraisal of the predictive models. In the case of Sn classes, they are expressed as:

$$\begin{aligned}
 \text{Precision Class 1 Sn} &= \frac{248}{272} = 0.912, \\
 \text{Recall Class 1 Sn} &= \frac{248}{254} = 0.976, \\
 \text{Precision Class 2 Sn} &= \frac{46}{53} = 0.868, \\
 \text{Recall Class 2 Sn} &= \frac{46}{69} = 0.667, \\
 \text{Precision Class 3 Sn} &= \frac{3}{5} = 0.600, \\
 \text{Recall Class 3 Sn} &= \frac{3}{7} = 0.429, \\
 \text{Precision Class 4 Sn} &= \frac{3}{3} = 1.000, \\
 \text{Recall Class 4 Sn} &= \frac{3}{3} = 1.000,
 \end{aligned}
 \tag{15}$$

**Table 7**  
Error rates for W predicted with Cluster, CART, MLP and SVM methodologies.

Output data	Input data	Classification problem						
		Cluster		CART	MLP	SVM		
		Training error rate	Test error rate	Error rate	Training error rate	Test error rate	Training error rate	Test error rate
W	Input N1	0.72	0.83	0.11	0.10	0.25	0.00	0.20
	Input N2	0.66	0.63	<b>0.09</b>	0.10	0.25	0.00	0.20
	Input N3	0.51	0.50	0.12	0.20	0.20	0.20	0.20
	Input N4	0.76	0.78	0.14	0.20	0.20	0.01	0.25
	Input N5	0.48	0.73	0.11	0.20	0.20	0.00	0.20
	Input N6	0.69	0.62	0.12	0.20	0.20	0.00	0.20

The best result is indicated in bold.

**Table 8**  
Confusion matrix of the best CART model for Sn prediction.

		Predicted Sn class				
		Class 1	Class 2	Class 3	Class 4	
Real Sn class	Class 1	248	5	1	0	Total real class 1 = 254
	Class 2	22	46	1	0	Total real class 2 = 69
	Class 3	2	2	3	0	Total real class 3 = 7
	Class 4	0	0	0	3	Total real class 4 = 3
		Total pred. class 1 = 272	Total pred. class 2 = 53	Total pred. class 3 = 5	Total pred. class 4 = 3	

**Table 9**  
Confusion matrix of the best CART model for W prediction.

		Predicted W class			
		Class 1	Class 2	Class 3	
Real W class	Class 1	254	9	1	Total real class 1 = 264
	Class 2	15	46	1	Total real class 2 = 62
	Class 3	2	4	1	Total real class 3 = 7
	Total pred. class 1 = 271	Total pred. class 2 = 59	Total pred. class 3 = 3		

and regarding the prediction of W classes as

$$\text{Precision Class 1 W} = \frac{254}{271} = 0.937,$$

$$\text{Recall Class 1 W} = \frac{254}{264} = 0.962,$$

$$\text{Precision Class 2 W} = \frac{46}{59} = 0.780,$$

$$\text{Recall Class 2 W} = \frac{46}{62} = 0.742,$$

$$\text{Precision Class 3 W} = \frac{1}{3} = 0.333,$$

$$\text{Recall Class 3 W} = \frac{1}{7} = 0.143. \quad (16)$$

Considering the obtained results, the overall performance of the model is good, predicting more accurately than previously those classes with a greater number of observations (Table 3).

Finally, Figs. 9 and 10 show the decision trees for Sn and W, respectively, which outperformed in the herein research work. A decision tree for Sn using Input N1 (error 0.10; Table 6), considers P, Be, As and Cr concentrations in first place for Sn prediction (Fig. 9). In the W decision tree using Input N2 (error 0.09; Table 7), the concentrations of Cd, Fe, Cr and As are considered the best elements for W prediction (Fig. 10).

In summary, the CART is selected as the best technique, considering:

1. Lower error rates: compared to the rest of the techniques, indicating a problem where non-linearities have a low impact;
2. Simplicity: The model showed high performance in different pruning rules scenarios;
3. The capacity for interpretation: CART provides the error rate output and establishes the priorities related with each of the prediction variables, thus assigning a hierarchy within the input dataset.

The couple distribution of estimated and predicted Sn and W stream sediment contents (Fig. 11) show high degrees of association. Almost higher Sn and W contents distribution are coincident with old Sn and W mines (Fig. 1b). However, local areas located in the central zone are enriched in these elements and could be suggested as future exploitation zones.

#### 4. Conclusions

Different classifiers were tested using a stream sediments dataset where trace element concentrations were used for prediction of a selected target element.

To build a more adequate and accurate predictive model for Sn and W concentrations, the algorithms – CART, MLP, SVM, clustering - were applied with six previously selected inputs (N1-N6) and compared as a regression or as a classification application.

The solutions obtained from the regression and classification problem indicate that CART achieves the best predictions both for Sn and W. In the example of the regression approach, coefficients of determination of 0.67 for Sn and 0.70 for W were obtained, while in the classification approach, an error rate of 0.10 was received in the prediction of both Sn and W.

In conclusion, the greatest accuracy of classification process using CART methodology suggests that it is the best prediction for Sn and W from stream sediments in Lardosa.

The proposed models to predict element concentrations are very useful, particularly for the identification of enriched mineralized areas and on the definition of future exploitation areas. However, the application of this methodology to new areas or conditions needs a new training stage so that the predictive models can adapt to the new data.

Moreover, such databases and expert systems can also be built for



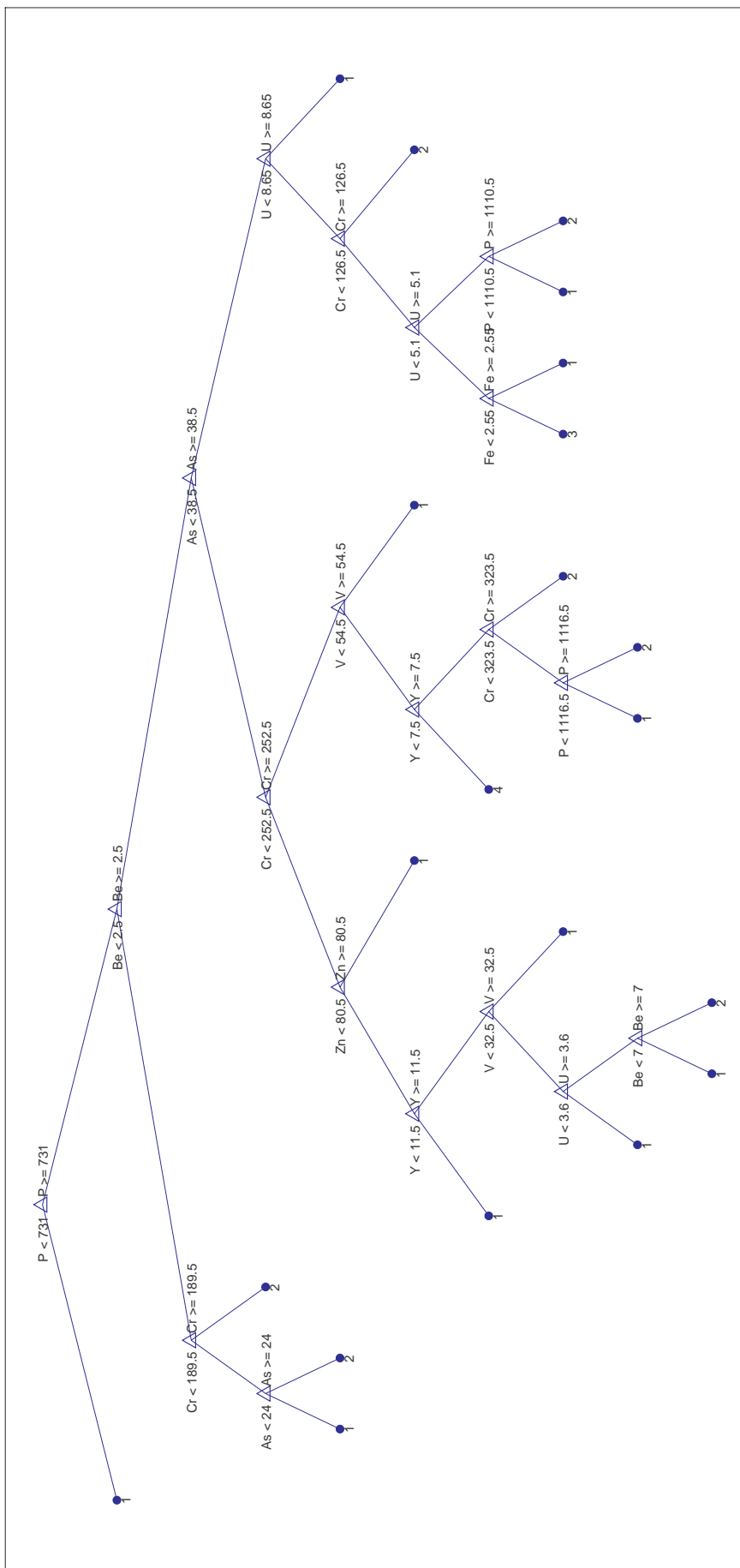


Fig. 9. Decision tree for Sn with Input N1 (error rate = 0.10).

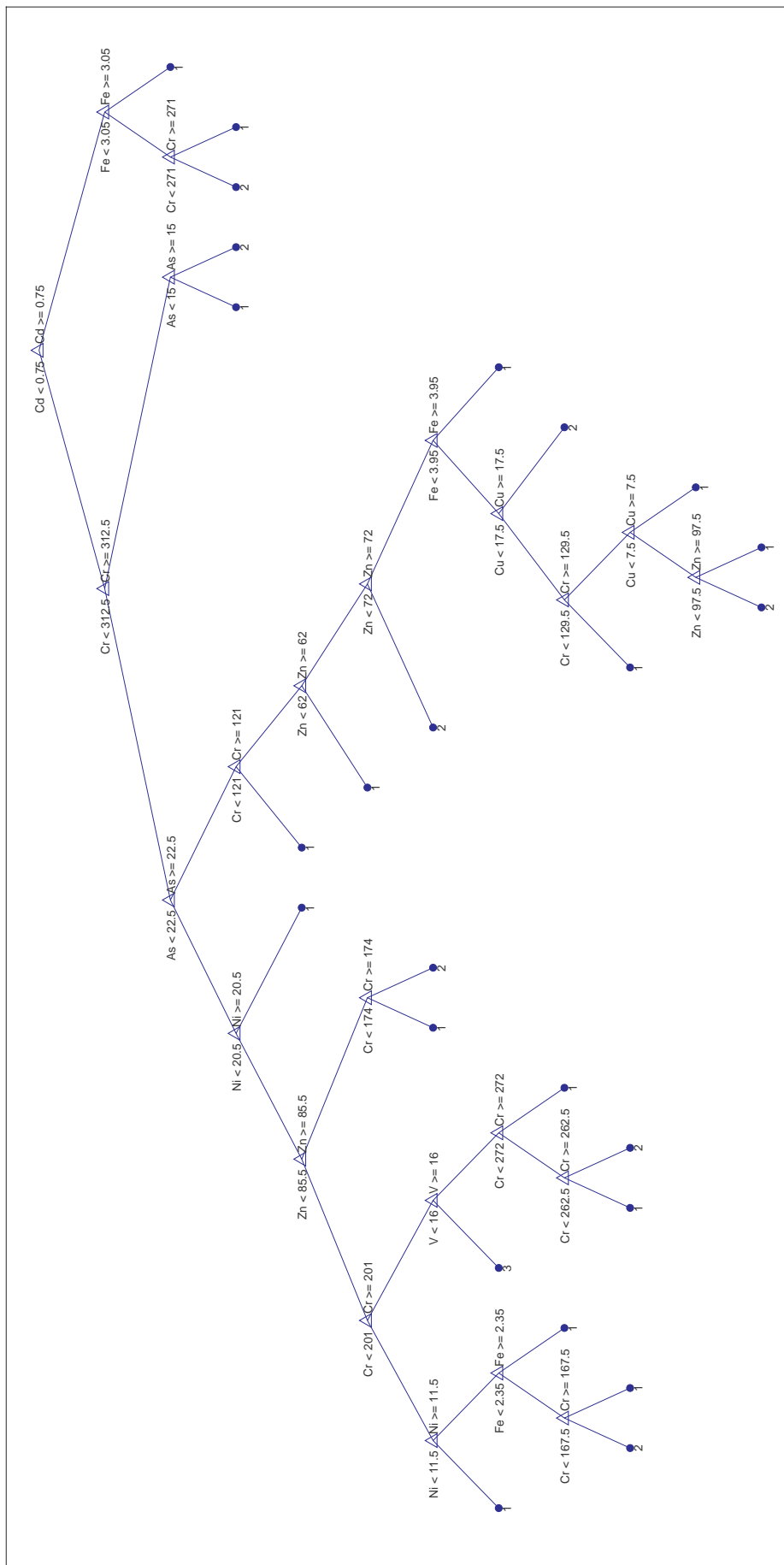


Fig. 10. Decision tree for W with Input N2 (error rate = 0.09).

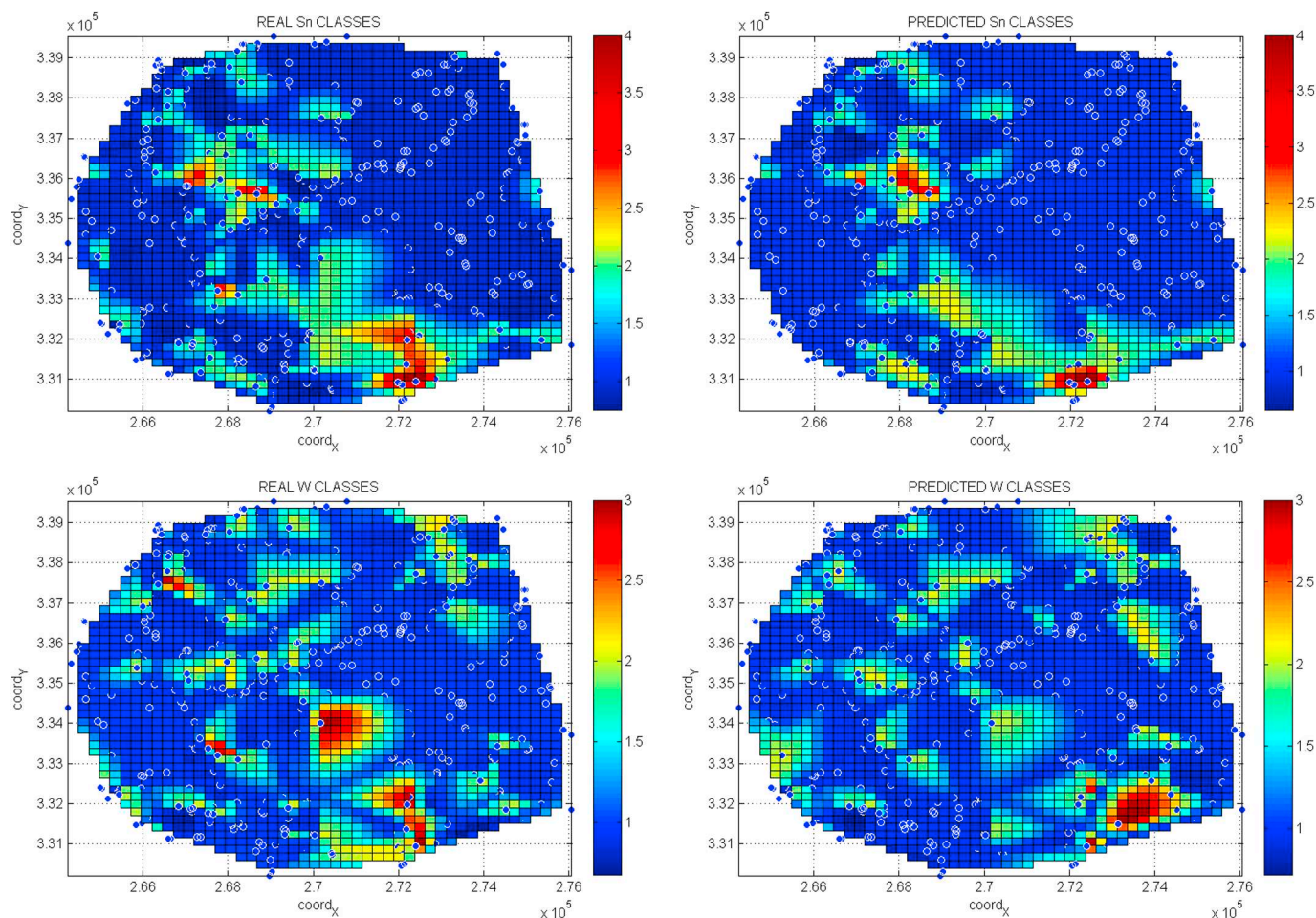


Fig. 11. Spatial distribution of real and predicted Sn and W classes from the studied area.

the routine monitoring of different potential contaminant substances in different media (Daszykowski et al., 2015).

In a future line of research, a stochastic spatial approach, through geostatistical techniques, will allow for the definition of concentration spatial patterns with spatial uncertainty incorporation and thus establishing coupled and empowered models for metal resources exploration.

#### Declaration of competing interest

The authors want to ensure that there are no conflicts of interest.

#### Acknowledgments

Thanks are due to Prof. M.R. Machado Leite for the use of data on stream sediments from Instituto Geológico e Mineiro, S. Mamede de Infesta (Portugal). C. Iglesias acknowledges the Spanish Ministry of Education, Culture and Sports for FPU 12/02283 grant. This research was carried out under the CERENA/FEUP (Natural resources and Environment Center), Portugal. The author acknowledges the funding provided by the Institute of Earth Sciences (ICT), under contracts UID/GEO/04683/2013 with FCT (the Portuguese Science and Technology Foundation) and COMPETE POCI-01-0145-FEDER-007690.

#### References

Antunes, I.M.H.R., Neiva, A.M.R., Silva, M.M.V.G., Corfu, F., 2009. The genesis of I- and S-type granitoid rocks of the Early Ordovician Oledo pluton, Central Iberian Zone (central Portugal). *Lithos* 111, 168–185. <https://doi.org/10.1016/j.lithos.2008.07.014>.

Antunes, I.M.H.R., Albuquerque, M.T.D., Sanches, F.A.N., 2014. Spatial risk assessment

related to abandoned mining activities: an environmental management tool. *Environ. Earth Sci.* 72, 2631–2641. <https://doi.org/10.1007/s12665-014-3170-4>.

Bahn, V., McGill, B.J., 2013. Testing the predictive performance of distribution models. *Oikos* 122, 321–331. <https://doi.org/10.1111/j.1600-0706.2012.00299.x>.

Ballabio, C., Sterlacchini, S., 2012. Support vector machines for landslide susceptibility mapping: the Staffora River Basin case study, Italy. *Math. Geosci.* 44, 47–70. <https://doi.org/10.1007/s11004-011-9379-9>.

Bennett, K.P., Mangasarian, O.L., 1992. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software* 1, 23–34. <https://doi.org/10.1080/10556789208805504>.

Bishop, C.M., 2008. *Neural Networks for Pattern Recognition*. Oxford University Press, New York (504 pp.).

Boots, B.N., Getis, A., 1988. *Point Pattern Analysis*. Sage Publications, Newbury Park, Calif (93 pp.).

Bowden, P., Jones, J., 1978. Mineralization in the younger granite province of northern Nigeria. *Met Assoc with Acid Magmat* 3, 179–190.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and Regression Trees*. Wadsworth Inc, Belmont, New York (368 pp.).

Brown, S.D., Myles, A.J., 2009. Decision tree modeling in classification. In: Brown, S.D., Tauter, R., Walczak, B. (Eds.), *Comprehensive Chemometrics*. Oxford-Elsevier, pp. 541–569.

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Disc.* 2, 121–167.

Carranza, E.J.M., 2011. Geocomputation of mineral exploration targets. *Comput. Geosci.* 37, 1907–1916. <https://doi.org/10.1016/j.cageo.2011.11.009>.

Chen, Y., Wu, W., 2017. Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data. *Geochemistry: Exploration, Environment, Analysis* 17, 231–238.

Chen, Y., Lu, L., Li, X., 2014. Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly. *J. Geochem. Explor.* 140, 56–63.

Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297. <https://doi.org/10.1007/BF00994018>.

Cristianini, N., Shawe-Taylor, J., 2000. *Support Vector Machines*. Cambridge University Press, Cambridge (204 pp.).

Daszykowski, M., Korzen, M., Krakowska, B., Fabianczyk, K., 2015. Expert system for monitoring the tributyltin content in inland water samples. *Chemom. Intell. Lab. Syst.* 149, 123–131. <https://doi.org/10.1016/j.chemolab.2015.10.008>.

- Diggle, P.J., 1983. Statistical analysis of spatial point patterns. In: *Mathematics in Biology*. 2 Academic Press, London-New York (690 pp.).
- Diggle, P.J., 1990. A point process modeling approach to raised incidence of a rare phenomenon in the vicinity of a pre-specified point. *Journal of Royal Statistical Society* 153, 349–362.
- Duan, P., Liu, C., Mo, X., Deng, J., Qin, J., Zhang, Y., Tian, S., 2018. Discriminating characters of ore-forming intrusions in the super-large Chalukou porphyry Mo deposit, NE China. *Geosciences Frontiers* 9 (5), 1417–1431.
- Falconer, J.D., 1912. Nigerian tin; its occurrence and origin. *Econ. Geol.* 7, 542–546. <https://doi.org/10.2113/gsecongeo>.
- Fan, H.-R., Yang, K.-F., Hu, F.-F., Liu, S., Wang, K.-Y., 2016. The giant Bayan Obo REE-Nb-Fe deposit, China: controversy and ore genesis. *Geosciences Frontiers* 7 (3), 335–344.
- Ferrier, S., 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Syst. Biol.* 51, 331–363. <https://doi.org/10.1080/10635150252899806>.
- Fletcher, R., 1989. *Practical Methods of Optimization*. John Wiley & Sons, Inc, New York (456 pp.).
- Freedman, D., Pisani, R., Purves, R., 2007. *Statistics*. W. W. Norton & Company, New York (697 pp.).
- Harris, P., Fotheringham, A.S., Crespo, R., Charlton, M., 2010. The use of geographically weighted regression for spatial prediction: an evaluation of models using simulated data sets. *Math. Geosci.* 42, 657–680. <https://doi.org/10.1007/s11004-010-9284-7>.
- Heaton, J., 2012. *Introduction to the Math of Neural Networks*. Heaton Research, New York (112 pp.).
- Ibrahim, A.M., Bennett, B., 2014. The assessment of machine learning model performance for predicting alluvial deposits distribution. *Procedia Computer Sciences* 36, 637–642. <https://doi.org/10.1016/j.procs.2014.09.067>.
- Iglesias, C., Martínez-Torres, J., García-Nieto, P.J., Alonso Fernández, J., Díaz, M.C., Piñeiro, J., Taboada, J., 2014. Turbidity prediction in a river basin by using artificial neural networks: a case study in Northern Spain. *Water Resources Management* 28, 319–331. <https://doi.org/10.1007/s11269-013-0487-9>.
- Instituto Geológico e Mineiro, 1998. *Reports From a Prospecting Project for Tungsten, Tin and Associated Minerals From Góis-Segura*. (Portugal).
- Iverson, L.R., Prasad, A.M., 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecol. Monogr.* 68 (4), 465–485. [https://doi.org/10.1890/0012-9615\(1998\)068\[0465:PAOTSF\]2.0.CO;2](https://doi.org/10.1890/0012-9615(1998)068[0465:PAOTSF]2.0.CO;2).
- Jackson, J.E., 1991. *A User's Guide to Principal Components*. John Wiley & Sons, New York (38 pp.).
- Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. *ACM Comput. Surv.* 31, 264–323.
- Kotlu, V., Deshpande, B., 2015. *Predictive Analytics and Data Mining*, first ed. Elsevier (446 pp.).
- Leite, E.P., de Souza Filho, C.R., 2009. Probabilistic neural networks applied to mineral potential mapping for platinum group elements in the Serra Leste region, Carajás Mineral Province, Brazil. *Comput. Geosci.* 35, 675–687. <https://doi.org/10.1016/j.cageo.2008.05.003>.
- Lek, S., Park, Y.S., 2008. Multilayer Perceptron. In: *Encyclopedia of Ecology*, pp. 2455–2462. <https://doi.org/10.1016/B978-008045405-4.00162-2>.
- McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biol.* 5, 115–133.
- Moisen, G.G., 2008. Classification and regression trees. In: *Encyclopedia of Ecology*, pp. 582–588. <https://doi.org/10.1016/B978-008045405-4.00149-X>.
- Oh, H., Lee, S., 2010. Application of artificial neural network for gold-silver deposits potential mapping: a case study of Korea. *Nat. Resour. Res.* 19 (2), 103–124. <https://doi.org/10.1007/s11053-010-9112-2>.
- Oommen, T., Misra, D., Twarakavi, N.K.C., Prakash, A., Sahoo, B., Bandopadhyay, S., 2008. An objective analysis of support vector machine based classification for remote sensing. *Math. Geosci.* 40, 409–424. <https://doi.org/10.1007/s11004-008-9156-6>.
- Porwal, A., Carranza, E.J.M., Hale, M., 2003. Artificial neural networks for mineral potential mapping: a case study from Aravalli Province, Western India. *Natural Resource Research* 12, 155–171. <https://doi.org/10.1023/A:1025171803637>.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* 71, 804–818. <https://doi.org/10.1016/j.oregeorev.2015.01.001>.
- Rowlingston, B.S., Diggle, P.J., 1991. Estimating the K-function for a Univariate Point Process on an Arbitrary Polygon. *Mathematics Department Technology Report, MA91/58*. Lancaster University (5 pp.).
- Santos Oliveira, J., Pedrosa, M., Canto, M.M., Rochas, S.J., 1998. Impacte ambiental provocado pela atividade mineira. Caracterização da situação junto da Mina de Jales, avaliação dos riscos e medidas de reabilitação. In: *Actas do V Congresso Nacional Geologia*. 84 (2), pp. 74–77.
- Scott, J.M., Csuti, B., 1997. *Gap Analysis for Biodiversity Survey and Maintenance*. Joseph Henry Press, Washington, DC.
- Skov, F., Svenning, J.C., 2004. Potential impact of climatic change on the distribution of forest herbs in Europe. *Ecography* 27 <https://doi.org/10.1111/j.0906-7590.2004.03823.x>. 366–38.
- Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. *Stat. Comput.* 14, 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>.
- Soman, K.P., Diwakar, S., Ajay, V., 2006. *Insight into Data Mining. Theory and Practice*. Prentice Hall of India, New Dehli (420 pp.).
- Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. *Math. Geol.* 34, 1–21. <https://doi.org/10.1023/A:1014009426274>.
- Thissen, U., Van Brakel, R., de Weijer, A., Melssen, W.J., Buydens, L.M.C., 2003. Using support vector machines for time series prediction. *Chemom. Intell. Lab. Syst.* 69 (1/2), 35–49. [https://doi.org/10.1016/S0169-7439\(03\)00111-4](https://doi.org/10.1016/S0169-7439(03)00111-4).
- Vapnik, V., 1982. *Estimation of Dependences Based on Empirical Data*. Springer Verlag, Berlin (505 pp.).
- Vapnik, V., 1998. *Statistical Learn Theory*. John Wiley and Sons, Inc, New York (768 pp.).
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemom. Intell. Lab. Syst.* 2, 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
- Zhang, C., Yu, X.-Y., Jiang, T.-L., 2018. Mineral association and graphite inclusions in nephrite jade from Liaoning, northeast China: implications for metamorphic conditions and ore genesis. *Geosciences Frontiers* 10 (2), 425–437. <https://doi.org/10.1016/j.gsf.2018.02.009>.
- Zhao, J., Chen, S., Zuo, R., 2016. Identifying geochemical anomalies associated with Au-Cu mineralization using multifractal and artificial neural network models in the Ningqiang district, Shaanxi, China. *J. Geochem. Explor.* 164, 54–64.
- Zuo, R., 2017. Machine learning of mineralization-related geochemical anomalies: a review of potential methods. *Nat. Resour. Res.* 26, 457–464.
- Zuo, R., Xiong, Y., 2018. Big data analytics of identifying geochemical anomalies supported by machine learning methods. *Nat. Resour. Res.* 27, 5–13.