Applied Geochemistry 123 (2020) 104779

Contents lists available at ScienceDirect

Applied Geochemistry

journal homepage: http://www.elsevier.com/locate/apgeochem

Hybrid multivariate typological model for the banded iron formations from the Bonito mine, Northeastern Brazil

Helano Regis da Nóbrega Fonteles [a,*], César Ulisses Vieira Veríssimo [b], Henrique Garcia Pereira [c], Irla Gonçalves Barbosa [d]

[a] Mineral Resources Specialist at Agência Nacional de Mineração (ANM), Avenida Padre Antônio Tomás, 2110, Aldeota, 60140-160, Fortaleza, CE, Brazil
[b] Full Professor at Departamento de Geologia, Universidade Federal do Ceará. Permanent address: Avenida Mister Hull, s/n. Bloco 913, 60455-760, Fortaleza, CE, Brazil
[c] Senior Researcher at CERENA (Centro de Recursos Naturais e Ambiente), Instituto Superior Técnico, Universidade de Lisboa, Avenida Rovisco Pais, 1, 1049-001, Lisboa, Portugal
[d] Departamento de Geologia, Universidade Federal do Ceará. Permanent address: Avenida Mister Hull, s/n. Bloco 913, 60455-760, Fortaleza, CE, Brazil

ARTICLE INFO

Handling editor. Prof. M. Kersten

Keywords:
Serra dos Quintos Formation
BIF typology
correspondence analysis
K-means clustering

ABSTRACT

Typological ore delineation is a basic procedure that provides various bases in ore geology on which crucial developments rely on, in particular, those supporting mineral resource evaluation. To deal with several complex issues related to typological delineation of heterogeneous ore bodies, an unusual hybrid multivariate approach, based on a skillful combination of geochemical/geological knowledge and correspondence analysis (CA), was established. This approach was tested on banded iron formations (BIF) in the Bonito mine, Northeastern Brazil, enhancing the available geochemical knowledge by revealing grade combinations related to geological processes causing such deposits. Petrographic studies helped identify four main BIF-types: amphibolitic, hematitic, martitic, and magnetitic itabirites. An exhaustive geochemical database comprising 1,384 BIF samples assayed for Fe_2O_3, SiO_2, Al_2O_3, P, and Mn grades was established and classified for the aforementioned BIF-types. The methodology provides an updated typological model based on a unique aspect of CA that allows a simultaneous projection of samples and their characteristics onto axes of the same abstract space; this helps validate mathematical results with geological information linked to sample attributes. The BIF typology model was established by applying K-means clustering to factorial scores derived through CA. Hence, the BIF typological model was updated, and new geochemical BIF-types helped interpret some of the main geological processes related to them. Geochemically, the factorial axes could be related to: 1) Iron rich BIF rocks, which are clearly represented on the positive side of Axis 1, comprising almost 50% of the total variance; 2) Manganese and alumina, associated with high silica content, by attachment to Axis 2, which describes a possible and critical terrigenous input.

1. Introduction

In the first decade of the 21st century, the quest for iron ore deposits was boosted by an increasing Chinese demand to feed its booming steel industry. Consequently, new economic criteria stemming from this scenario were developed to evaluate the feasibility of minor deposits, such as the Bonito mine, whose geological significance to the Neoproterozoic evolution of the Seridó Belt during the Brasiliano-Pan African orogeny is also addressed.

The Bonito iron ore mine, was originally divided into two groups based on varying Fe_2O_3 grades, as well as on their geochemical and petrological compositions (Barbosa, 2013). High-grade ores are represented by magnetitic ore (91.60-94.20% Fe_2O_3) and magnesian skarn (60.30-72.33% Fe_2O_3). The other iron ore group, considered here, comprises banded iron formations (BIFs) containing low-grade iron ores (30.97-60.30% Fe_2O_3), in contrast with the first group.

This study aims to better understand the geology of the Serra dos Quintos Formation, a metavolcanosedimentary sequence, by identifying BIF-types in the Bonito iron mine, in Jucurutu, Rio Grande do Norte State, Northeastern Brazil. The combined multivariate methodology presented here embraces the multi-dimensionality of the available geochemical database.

* Corresponding author.
E-mail addresses: helano.fonteles@anm.gov.br (H.R.N. Fonteles), verissimo@ufc.br (C.U.V. Veríssimo), henrique.pereira@ist.utl.pt (H.G. Pereira), irlag_barbosa@hotmail.com (I.G. Barbosa).

https://doi.org/10.1016/j.apgeochem.2020.104779
Received 30 March 2020; Received in revised form 20 August 2020; Accepted 21 August 2020
Available online 2 October 2020

For this purpose, the methodology applied to this deposit considers two aspects:

(a) An enhanced geological/geochemical study of the ore body was based on similarities/distinctions between formations, providing a conceptual model focused on their relationships. The 'nderlying data model may be viewed as an "n ' p matrix", with n rows representing 1,384 samples and p columns representing concentrations of Fe_2O_3, SiO_2, Al_2O_3, P, and Mn;
(b) The second aspect aimed at broadly classifying ore types with geochemical meaning, based on correspondence analysis as a purely geometric method to compress data in a "$p_1 < p$" artificial space, with samples and concentrations being simultaneously projected, using the chi-square distance applied to the matrix elements, appropriately classified into various grades.

We aim to put forward a new approach based on CA along with geological interpretation, stemming from the modeling of lithogeochemical data and petrographic BIF characterization. Based on an integrated mathematical/geochemical model, and by applying of a nonhierarchical partitional algorithm (K-means method), homogeneous (or acceptable heterogeneous) groups of samples were delineated, which are the elements of the required typology.The first consistent usage of automatic classification methodologies in geological sciences was performed by Parks, 1966 and – among others – by Rhodes, 1969, applying specifically to granites. CA, as put forward by Benzécri, 1973, 1977, Lebart et al., 1984, Mellinger, 1987, and applied by Pereira, 1981, is a classical distribution-free multivariate method. Being purely geometric, this method can help construct a unique artificial space where individuals (samples) and properties (variables) are jointly projected. Meaningful BIF-types are obtained when sample projections onto significant CA axes are clustered using a nonhierarchical method, which are plausibly explained in scientific terms taking advantage of previous CA/geochemical interpretations. Moreover, in the considered case, where the number of samples is relatively important (at the usual scale for orebody recognition), hierarchical classification cluster methods could be particularly cumbersome, leading to uninterpretable graphs due to the copious number of samples.

2. Geological setting

Regional framework. The Bonito mine is located in the Rio Piranhas-Seridó Domain (PSD) occupying an extensive area in the center portion of the State of Rio Grande do Norte, bounded to the east by the São José do Campestre Domain (SJD) and to the west by the Jaguaribeano Domain (JD). Both contacts between the domains are tectonic and defined by ductile shear zones of Porto Alegre (to the east) and Picuí-João Câmara (to the west), as shown in Fig. 1A (Angelim et al., 2006).

Among the several stratigraphic schemes proposed for the Seridó Group, the one adopted here was presented by Angelim et al., 2006. The Seridó Group consists of a metavolcanosedimentary sequence called the Serra dos Quintos Formation at its base (Sá and Salim, 1980; Ferreira and Santos, 2000). The Seridó Group is divided from the base to top as follows: (a) Serra do Quintos Formation (NPsq), comprising ferruginous quartzite, hematitic and/or magnetitic itabirite, garnet-tremolite schist, muscovite-quartzite, gneiss, actinolite-schist with magnetite and, occasionally, metaultramafic/metamafic rocks and leucogneiss; (b) Jucurutu Formation (NPj), mainly comprising biotite ± epitote ± amphibole paragneiss, with intercalations of a basal conglomerate, marbles, calc-silicate rocks, mica-schist, quartzite, iron formations, metavolcanic rocks, and metacherts; (c) Equador Formation, predominantly comprising muscovite-quartzite with arcosean facies containing interbedded polimitic metaconglomerate; and (d) Seridó Formation, comprising feldspathic mica-schist or aluminous medium to high-grade metamorphic facies across most of the unit. Locally, marble, calc-silicate rocks, quartzite, and metavolcanic rocks are interbedded within the main lithotype.

The Seridó Group was affected by a strong transpressional deformation, by granitic magmatism and reworking by transcurrent shear zones during the Brasiliano-Pan-African orogeny, ca. 575 Ma (Sá et al., 1995; Archanjo et al., 2013). This deformation is recorded throughout the Bonito mine by a penetrating tectono-metamorphic foliation in a dominant northeast direction, plunging preferentially toward the southeast, causing a large antiform with an N-S axial plane and south-dipping axis (Fig. 1A).Iron formations associated with marble, calc-silicate paragneiss, and quartzite of the Seridó Group could be part of a quartzite-pelite-carbonate sequence deposited in a shallow marine environment (Sá, 1994). The BIFs studied here were assumed to be a part of Serra dos Quintos Formation. Only the basal units of the Seridó Group (Serra dos Quintos and Jucurutu Formations) occur in the study area (Figs 1 and 2).Geochemical and isotopic studies conducted by Sial et al., 2015 for marble and BIFs of the Seridó Group, including a few samples from the Bonito mine, suggested a rift-type model with narrow basins and strong hydrothermal input, as previously proposed by Van Schmus et al., 2003 and Hollanda et al., 2015. Positive Eu e Ce anomalies and high Cr concentration could be explained by hydrothermal venting systems related to rifting with the creation of anoxic ferruginous conditions and leaching of mafic/ultramafic rocks (Sial et al., 2015).

Petrography of the Bonito mine BIFs. Barbosa, 2013 considered the BIFs (itabirites) of the study area to be metavolcanic sediments with alternating bands of silicate minerals (mainly amphiboles and quartz) and iron oxides (hematite, magnetite, martite, and goethite) that were locally affected by fluids with low redox potential, marked by the presence of pyrite and chalcopyrite. BIF-types could not be derived exclusively by exhaustively assessing available information. In fact, petrographic studies relying on both reflected and transmitted light microscopy have demonstrated their indispensability in establishing meaningful ore typology. Hence, a mathematical classification method combining all available information (such as CA) is crucial in dealing with site-specific typological issue. Based on mineralogy, texture, intergrain contacts, and modal percentage, itabirites include the following categories: 1) silicates; composed mainly of quartz, amphiboles (hornblende and tremolite/Fe-actinolite), and in lesser proportions, iron oxide and sulfides; and 2) iron oxides (hematite, magnetite, and martite), which are present in the samples and silicate minerals as well. Pyrite and chalcopyrite have also been described in some samples. The *amphibolitic itabirites (AmI)* are composed of tremolite/Fe-actinolite (15–89%), hornblende (20–70%), quartz (25–75%), magnetite (1–10%), and pyrite (1–10%). Their texture varies from grano-lepidoblastic (hornblende-amphibolitic itabirites) to lepido-granoblastic (tremolite/Fe-actinolite-amphibolitic itabirites). Under crossed-polars, quartz crystals are extensively anhedral showing shadowy extinction. Magnetite-martite are observed in 1–10% proportion, while pyrite is rare (Fig. 3b,f; Fig. 5c,d). *Hematitic itabirites (HI)* comprise quartz (50–60%), Fe-actinolite (15–25%), and hematite (20%). Magnetite/martite (1–4%) and pyrite (1%) may occur in low amounts. Their texture is predominantly lepidoblastic with inequigranular Fe-actinolite crystals exhibiting orientated lamellar shapes within the metamorphic foliation. Goethite appears to be an alteration product of this mineral. Hematite shows a highly reflective white-gray color. Magnetite was recognized in the polished section as having a rose-tint color and prismatic form. Pyrite crystals were identified as strong yellow-colored and small prismatic grains (Fig. 3e; Fig. 5a,b). *Magnetitic itabirites (MgI)* were the most representative petrographic BIF-type among the samples. The slightly banded texture varies from lepido-granoblastic to lepidoblastic. The main minerals are quartz (1–70%), amphiboles (10–30%, hornblende and/or tremolite/Fe-actinolite), magnetite (20–70%), and martite (1–50%). Pyrite and chalcopyrite occur in low amounts (1–5%). In some samples in which hornblende was identified as the dominant amphibole, epidotization was observed as an alteration product (Fig. 3c; Fig. 4e,f; Fig. 5e,f). Tremolite/Fe-actinolite crystals usually show

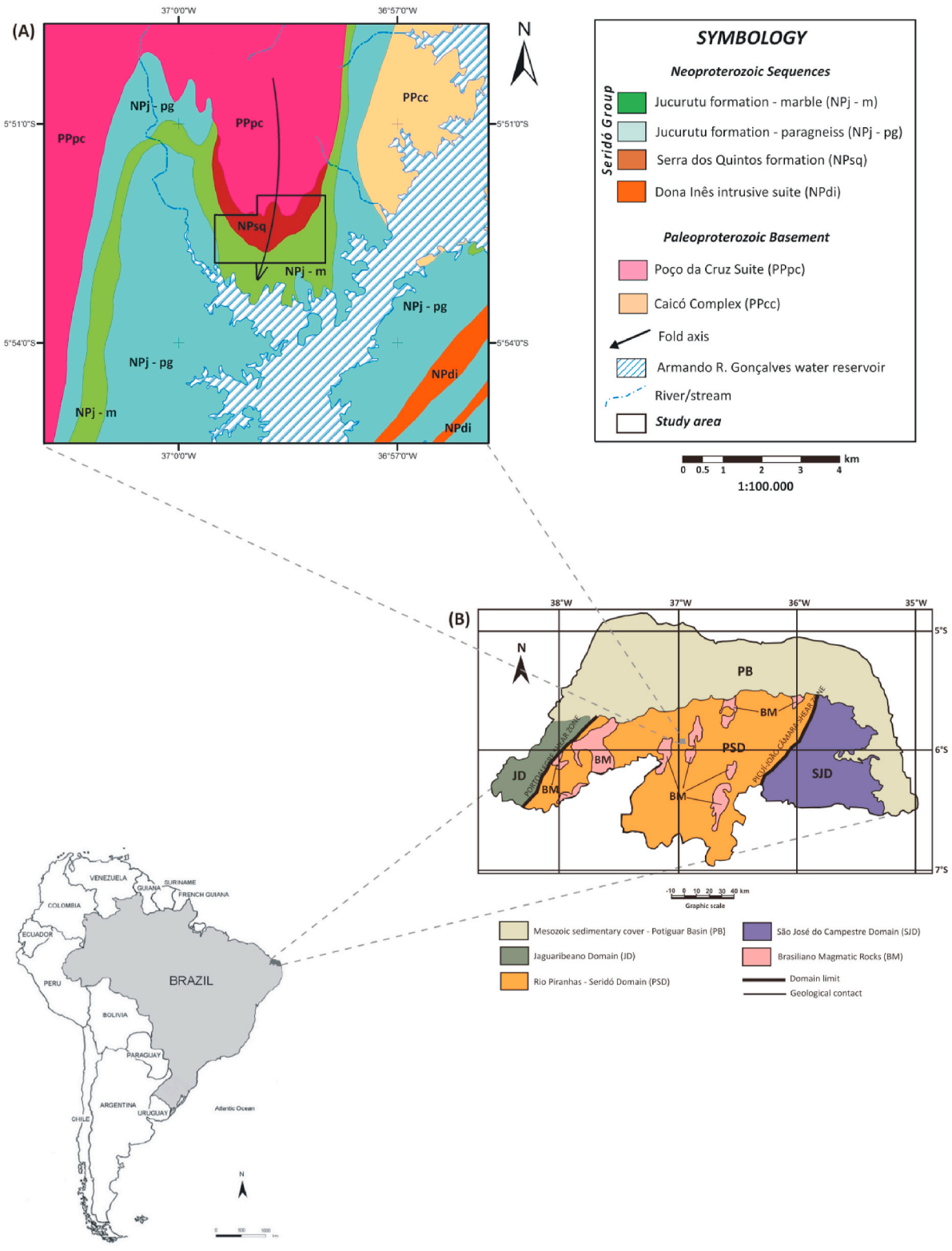

Fig. 1. Locational and simplified regional geological map. **A)** Regional geological framework. **B)** Geotectonic sketch map showing critical units and shear zones (Adapted after Angelim et al., 2006).

lamellar/acicular forms and strong pleochroism under crossed-polars. On polished sections, pyrite crystals are easily identified due to their highly reflective yellow color and cubic hipidiomorphic form. Chalcopyrite (most reflective yellow color) was observed as mineral inclusions on pyrite crystals (Fig. 3a; Fig. 4a,b). *Martitic itabirites (MI)* have specific petrographic features in a genetic and evolutionary context of the Bonito mine BIFs. On polished sections, martitization was interpreted as a mineral-phase transformation enhanced by substituting primary magnetite with martite, – a pseudomorph of primary iron ore (Ramdohr, 1981; Craig and Vaughan, 1984). A lepido-granoblastic metamorphic

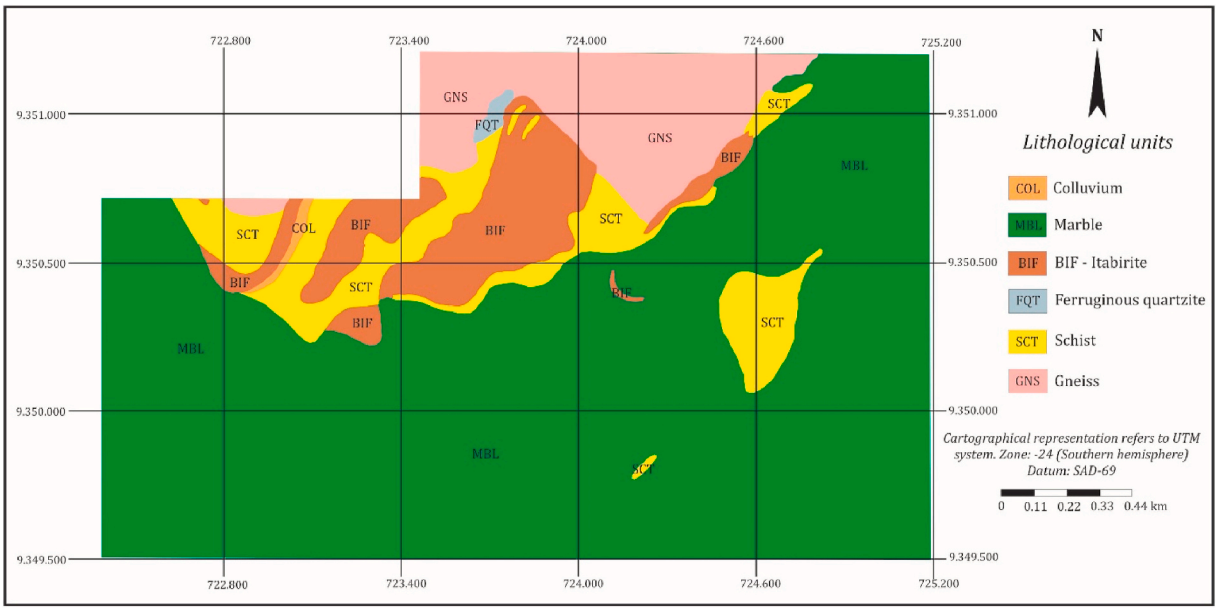

Fig. 2. Local geology map displaying the main lithological units at the research site (Adapted and simplified after MHAG, 2013).

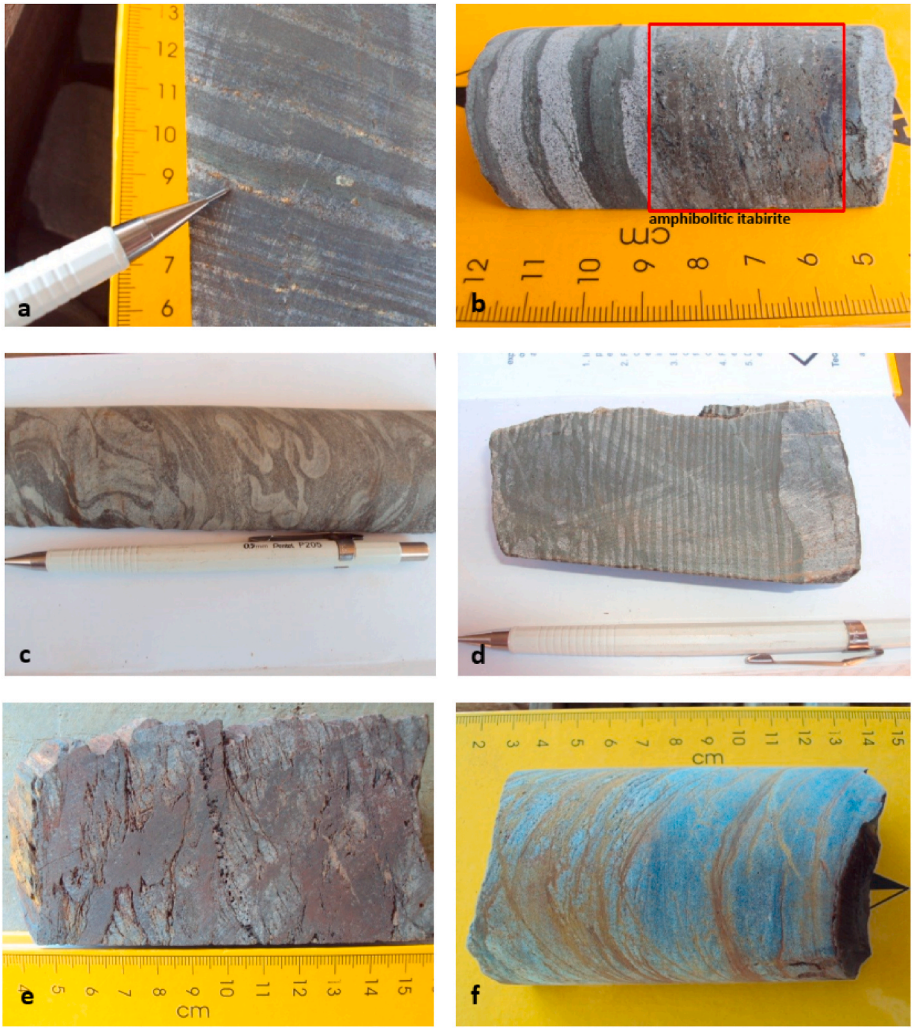

Fig. 3. Samples from the borehole-cores showing some of the main mineralogical and structural features of the Bonito mine low-grade BIFs: **a)** Magnetitic itabirite with millimetric sulfide-veins (pyrite); **b)** Magnetitic/martitic itabirite associated with amphibolitic itabirite; **c)** Magnetitic itabirite showing local intrafoliation folds; **d)** Martitic itabirite with an almost rhythmic banding; **e)** Irregular banding on hematitic itabirite with ore pods; and **f)** Fractured amphibolitic itabirite filled with goethite.

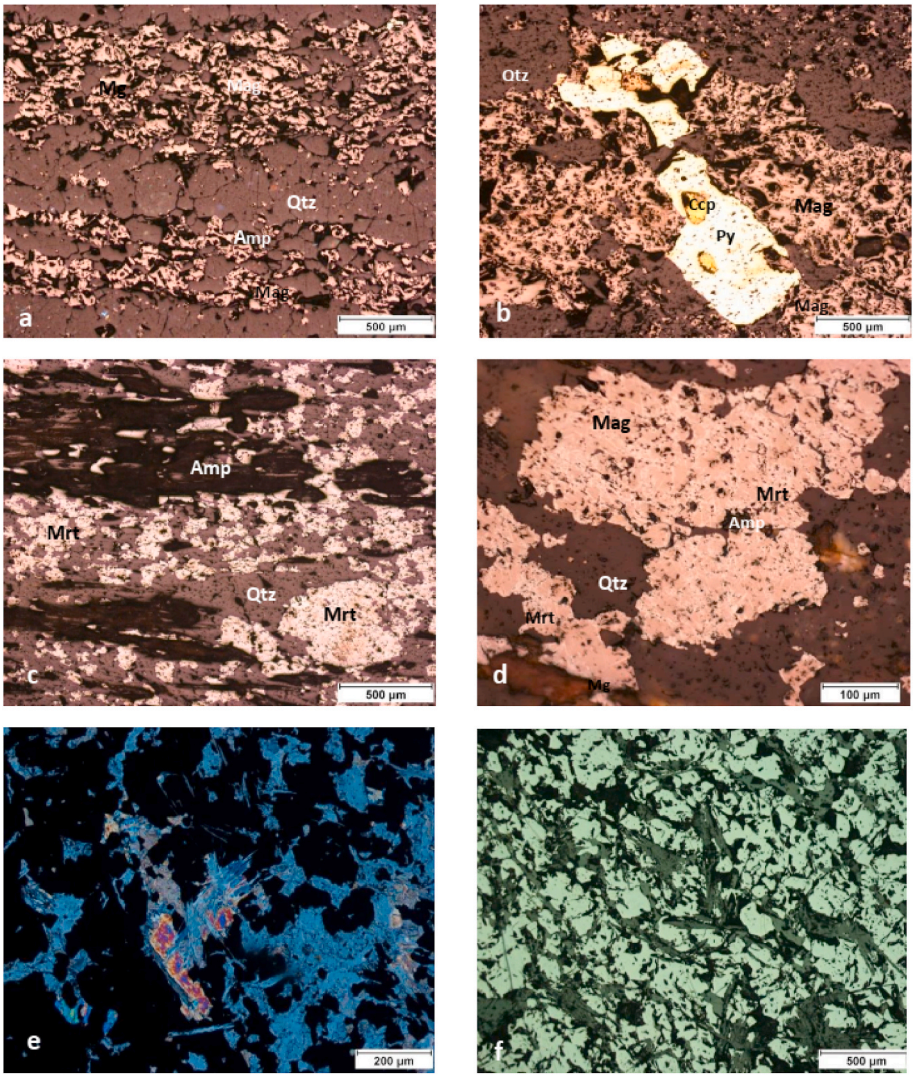

Fig. 4. Photomicrographs of the studied itabirites: **a)** Magnetitic itabirite with well-developed bands of quartz and magnetite+hornblende. Magnetite has a subhedral cubic form; **b)** Pyrite crystals exhibit cubic (yellow) with anhedral chalcopyrite (dark-yellow). These sulfides could occasionally include minerals from the itabiritic fabric; **c)** Martitic itabirite. Martite crystals preserving the habit of complete pseudomorphosed magnetite; **d)** Magnetitic itabirite partially martitized; **e)** Magnetitic itabirite on the thin-section with crossed-polars showing anhedral crystals of tremolite/Fe-actinolite with strong pleochroism and fibrous habit; and **f)** Polished section of magnetitic itabirite with highly reflective subhedral magnetite crystals under plane-polarized light. Abbreviations: quartz (Qtz), amphibole (Amp), magnetite (Mag), martite (Mrt), pyrite (Py), and chalcopyrite (Ccp).

texture is marked by association with quartz (30–60%), amphiboles (20–25%), martite (15–35%), magnetite (2–15%) and pyrite (<2%) (Fig. 3d; Fig. 4c,d). Natural magnetism was preserved, as it was observed in drill-core samples.

3. Materials and methods

3.1. The database

The geochemical database exploited here was obtained from an exhaustive recognition program developed to support economic feasibility studies for the Bonito iron ore mine by MHAG Serviços e Mineração company, owner of the mineral rights.

This database was structured to comprise quantitative and qualitative elements: geological records concerning lithology, borehole-intervals, geotechnical data, drilling-orientation paths, and sets of samples for geochemical analyses.

The samples were prepared at a geomechanics laboratory of MHAG and sent to SGS-GEOSOL Labs. Ltd. (Minas Gerais State, Brazil) for treatment with lithium tetraborate fusion and XRF spectrometry to determine Fe_2O_3, SiO_2, Al_2O_3, Cr_2O_3, V_2O_5, TiO_2, CaO, MgO, K_2O, Na_2O, P, and Mn concentrations. The loss on ignition (LOI) of the samples was analyzed on by calcination at 1000 °C until they attained constant mass.

BIF samples were collected for a petrographic investigation based on 26 polished sections and 12 thin-sections. The quantitative variables considered for the CA included concentrations of Fe_2O_3, SiO_2, Al_2O_3, P, and Mn. Originally, the database comprised all lithological records of 127 boreholes and included non-BIF lithologies, e.g., granite, schist, marble, and gneiss; however only BIF borehole records were considered for the first round of data analysis to focus on BIF-types. Considering the entire database, an emphasis was given to low-grade BIFs with $Fe_2O_3/SiO_2 \leq 1.00$.

Other major oxides were excluded from the multivariate analysis as not all BIF samples had the required input values. The geological record contained 2,557 lithological descriptions with/without geochemical data; however downsizing the working database to 1,384 samples was imperative for an objective analysis.

ANDAD 7.20 (CVRM, 2002) was applied in entirety for all factorial analyses. ESRI ArcMap® GIS 10.1 and Statistica® 10 (Statsoft Inc, 2010) were used for spatial data management and K-means clustering.

3.2. Multivariate modeling tools

3.2.1. Correspondence analysis (CA)

It is a geometric data treatment methodology focused on qualitative variables aimed at presenting tabular data graphically to assist in interpretation. The basic idea behind the methodology developed in the 1960s by Jean-Paul Benzécri (Benzécri, 1973, 1977) is that any matrix (input table) of non-negative numbers put together in a series of contingency tables could be converted into a series of two-dimensional plots

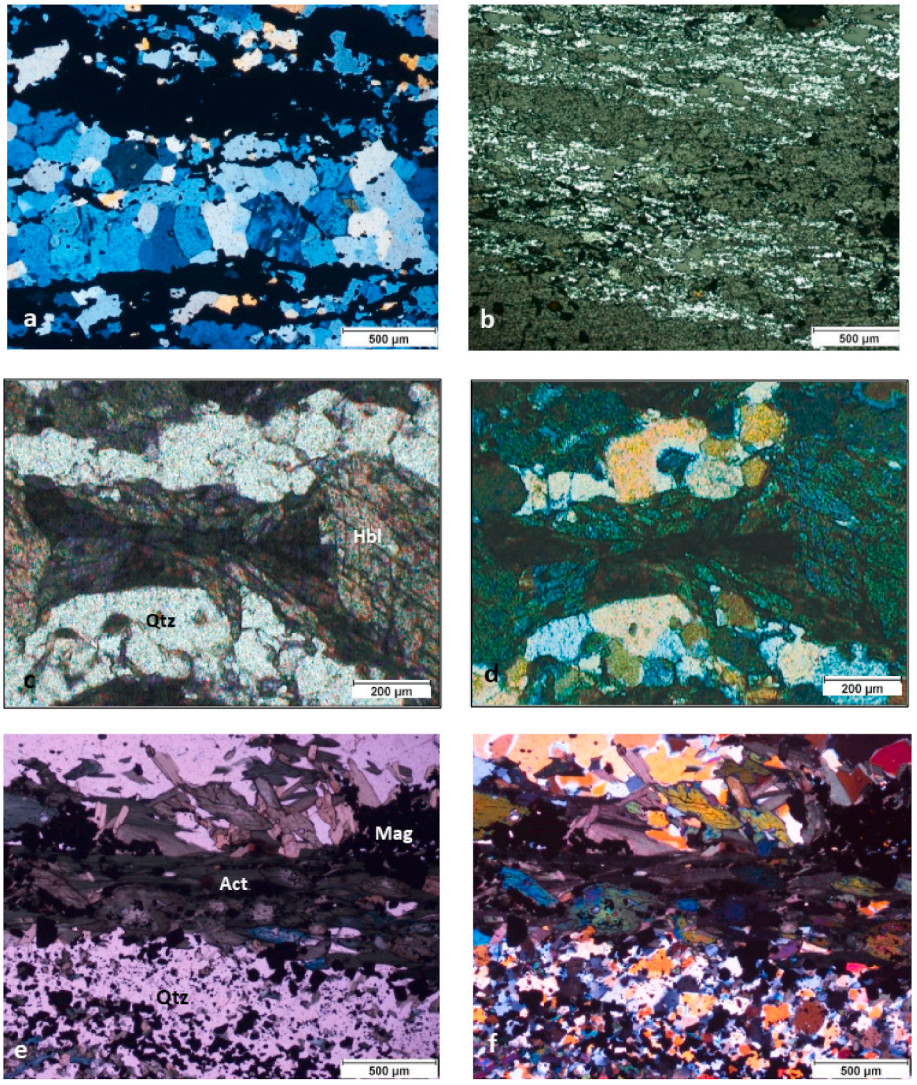

Fig. 5. Photomicrographs of the studied itabirites: **a)** Thin-section on crossed-polars. Hematitic itabirite with typical banding and a grano-lepidoblastic texture; **b)** Same sample. Polished section showing anhedral porous hematite crystals with highly reflective white-coloration. Pyrite crystals occurring in a 1% proportion; **c)** Amphibolitic itabirite. Thin-section on plane-polarized light where a hornblende crystal can be seen having typical cleavage-planes, positive relief, and strong greenish-brown coloration and a grano-lepidoblastic texture; **d)** Same sample on crossed-polars. Anhedral quartz crystals can be identified due to shadowy extinction; **e)** Silicate band from a sample of magnetitic itabirite (plane-polarized light); and **f)** Same sample under crossed-polars. Abbreviations: Actinolite (Act), magnetite (Mag), quartz (Qtz), and hornblende (Hbl).

representing rows and columns for items in the same graph.

A factorial analysis primarily aims to minimize multidimensionality of large datasets (Hill, 1974; Lebart et al., 1984; Mellinger, 1987; Grenacre and Blasius, 2006). Avoiding unnecessary a priori assumptions as far as possible, CA helps reveal and evaluate relationship patterns between input data (previously unintelligible in tabular form).

The mathematical rules for interpretation arise directly from the CA algorithm, as put forward by Benzécri. The first requirement, before applying the algorithm, is to ensure that the input matrix is a valid concatenation of contingency tables cross-tabulating two qualitative variables, with the modalities of the variables in each contingency table encoded so as to ensure comparability and unambiguity.

For grade information of the Bonito mine, the above-mentioned requirement was met, as the initial matrix by crossing samples x grades can be viewed as a cross-tabulation of two "qualitative variables" wherein rows are modalities of the variable *sample* and columns are modalities of the variable *grade*. Hence, we can sum up grades for a given sample, adding up to 100 if the grade is given in percentage and along rows, for one column, representing the *quantity of metal* (divided by 100, as the grade is given in percentage) for all samples (in the case of Fe and Mn), and dividing the total weight of SiO_2, Al_2O_3, and P in all samples by 100. A similar rationale applies when contingency tables are concatenated in several ways, based on the premise that the closure property is always met.

Unlike other factorial methods, CA is not affected by closure issues owing to some situations concerning geochemical concentrations. Multivariate modeling with CA is based on encoding original geochemical variables as distributional intervals translated into new categorical variables (Mellinger, 1987).

As eigenvalues are calculated through diagonalization of the matrix of the column profiles, the individuals (samples) on the data matrix are treated as conditional probabilities for each column. Thus, variables and samples are simultaneously displayed in the same factorial space due to the χ^2 (chi-square) metric, which is used to compute distances among samples. Hence, the independence of the variables is guaranteed by the mathematical constraints of the method (Teil, 1975; Mellinger, 1987; Grenacre and Blasius, 2006).

A useful advantage of CA over other more frequently applied multivariate factorial statistical methods, e.g., Principal components analysis (PCA) is that (in the most usual case of $n>p$) such a method always produces a "p axis" (in the most usual case where proportionality between columns is lacking), whereas CA always produces "p-1 axes" for each concatenated contingency table (the dimensional reduction begins on a smaller basis).The interpretation as per the scope of the CA, as put forward by Benzécri, 1973, 1977, Lebart et al., 1984, and Pereira et al., 2015, suggest that relationships between all projections are disclosed in terms of their linkage to each axis. Once outlined as per the first interpretation, there is usually a need to modify the encoding of the variables

(and sometimes, of the individuals) in order to achieve suitable results. Any improvement by interactive encoding is considered "satisfactory" when the model emerging through CA satisfies the relevant geological/geochemical information.

Furthermore, in line with the CA theory, the model ultimately obtained as a result of the methodology is not "validated" by any statistical hypothesis, but by its ability to yield valuable and helpful insights, also helping enhance a priori geological/geochemical knowledge.Comprehensive theoretical explanations and several geological applications have been discussed well by Balladur, 1970; Benzécri, 1973, 1977; Hill, 1974; Teil, 1975; Teil and Cheminee, 1975; Teillard and Volle, 1976; Valenchon, 1982; Lebart et al., 1984; Mellinger, 1987 ; Birks, 1987; Reis et al., 2004; Grenacre and Blasius, 2006; Patinha et al., 2007 and, Pereira et al., 2015.

3.2.2. K-means clustering method

This class of partitioning methods refers to gathering objects into mathematically similar groups around a point or centroid considered representative of a group. The groups are generated as a broad initial partition with no imposed hierarchical structure. The mathematical rules for creating the groups account for maximum similarities between objects within a group and maximum variance between groups.

The K-means clustering method, among others, is frequently used because of its simplicity, ease of implementation, efficiency, and acceptable results under low computational costs (Jain, 2010; Amorim, 2016). The common thread between nonhierarchical grouping methods is that the final scheme is built as a non-supervised classification model.

According to Jain, 2010, the algorithm is described by the following formula: let $X = \{x_i\}$, where i = 1,...,*n*, for a set of *n* *d*--dimensional points to be grouped into a set of *K* clusters, $C = \{c_k , k = 1,...,K\}$ in the Euclidean space. The K-means routine must find a partition such that the squared mean error between the empirical mean (μ_k) of a cluster (c_k) and the point of the cluster will be minimized given by

$$J\left(c_k\right) = \sum_{x_i \in c_k} \left| |x_i - \mu_k| \right|^2 \quad (1)$$

Eqn 1 defines the minimum variance between objects in a cluster, i.e., they have maximum similarity. The empirical mean (μ_k) is interpreted as the centroid of the cluster (c_k). The general solution of K-means clustering is obtained by minimizing the sum of the squared mean errors given by

$$J(C) = \sum_{k=1}^{K} \sum_{x_i \in c_k} \left| |x_i - \mu_k| \right|^2 \quad (2)$$

K-means clustering is fundamentally executed recursively until convergence and/or until it reaches the final number of pre-defined K clusters according to Eqn 1 and Eqn 2. Maximizing variance between clusters means that the dissimilarity between K groups is mathematically the highest.

In most cases, arbitrary seeds or centroids are assigned as starting points in the iterative process (Jain, 2010), which are iteratively updated until their positions meet the constraints of Eqn 1 and Eqn 2. The outcomes are represented by clusters with their centroids gathering a finite number of objects.

Similar to hierarchical clustering methods, K-means clustering also calls for a stopping-rule based on a mathematical criterion that could help reveal the "ideal" number of clusters. A dendrogram helps intuitively interpret its graphical display, while the computation of cluster centers only shows samples gathered around each center.

Some studies throughout the decades dealt with stopping-rules or clustering indexes. Among these, Calinski and Harabasz, 1974 presented a criterion based on variance within/between groups; Milligan and Cooper, 1985 evaluated the performance of 30 clustering criteria under several conditions through a series of Monte Carlo simulations; Rousseeuw, 1987 proposed a measure based on the average dissimilar distances between points within a cluster and the average minimum distance between these points and other clusters; Bezdek and Pal, 1998 reviewed two clustering methods (K-means and single linkage) through evaluation of three validation indexes and examined their most reliable technical features; Saitta et al., 2008 and Liu et al., 2010 assessed some of the well-known validation indexes and presented validation measures for the best classification of a set of data points.

4. Results of multivariate data modeling

4.1. Exploratory analysis

Among 127 drilling boreholes, 86 represented BIFs on superficial and/or sub-superficial levels. Therefore, these boreholes provided the main geological information (Fig. 6).

The entire dataset is represented by 1,384 BIF samples assayed for five grades (Fe_2O_3, SiO_2, Al_2O_3, P, and Mn), which are the variables arranged as columns of the basic matrix. Table 1 lists the basic descriptive statistics. Note that high CV values could signify high-variability.

4.2. BIF typology model based on a multivariate hybrid approach

Geochemical evaluation using CA. A new geometric-space was obtained using CA, where samples and variables were jointly projected. In contrast to some algorithms such as cluster/discriminant analysis, CA demands an intrinsic knowledge of the underlying geochemical phenomena for a reasonable interpretation of the results, projections onto axes defining the new geometric-space.

As the CA method requires qualitative variables as input, the grade-range for each variable was split into two/three classes. The limits of each class were iteratively established so that each class conveyed a geological meaning; classes were ultimately split (Table 2). Hence, each initial quantitative variable (grade) was substituted for a code (*1*, if the grade belonged to a prescribed class, and *0* otherwise). Thus, the initial table was transformed into an indicator matrix.

Table 3 shows the eigenvalues calculated through iterations considering values presented in Table 2. At first sight, one can observe that 79.11% of data variance is explained by the first three axes, even though 66.33% clearly describes the main scenario where the class-encoded variables help build an idea for the typological model. It is noteworthing that seven axes are sufficient to explain a total of 12 classes, given that one degree of freedom is subtracted for each one of the five contingency tables whose concatenation gave rise to the input indicator matrix.

Table 4 reveals the main absolute contributions spread along the seven axes representing the influence of each class on an axis. The variables Fe1 and Fe3 together contribute up to 28% of on Axis 1. The same can be noted with Al1 and Al2 which are well-positioned on Axis 2 (Fig. 7).

Here, the use of CA as a modeling-tool required that some samples be studied petrographically. Some geochemical patterns could be derived by interpreting factorial axes; although, without some guiding samples for a supervised classification no meaningful geological information could not be extracted and/or combined with other available data (Fig. 8). Therefore, this crucial information was provided by the petrographic and geochemical data (Table 5).

Application of K-means clustering: building a bridge between the factorial space and BIF typological model. K-means clustering was applied to establish the limits between BIF-types first interpreted as domains on the factorial space. A numerical solution was obtained through an algorithm using Statistica® 10 with the original code developed by Hartigan and Wong, 1979. For this task, the input data was conveyed by factorial coordinates of each sample from axes 1 and 2 (Fig. 8). K-means clustering was conducted considering that the first

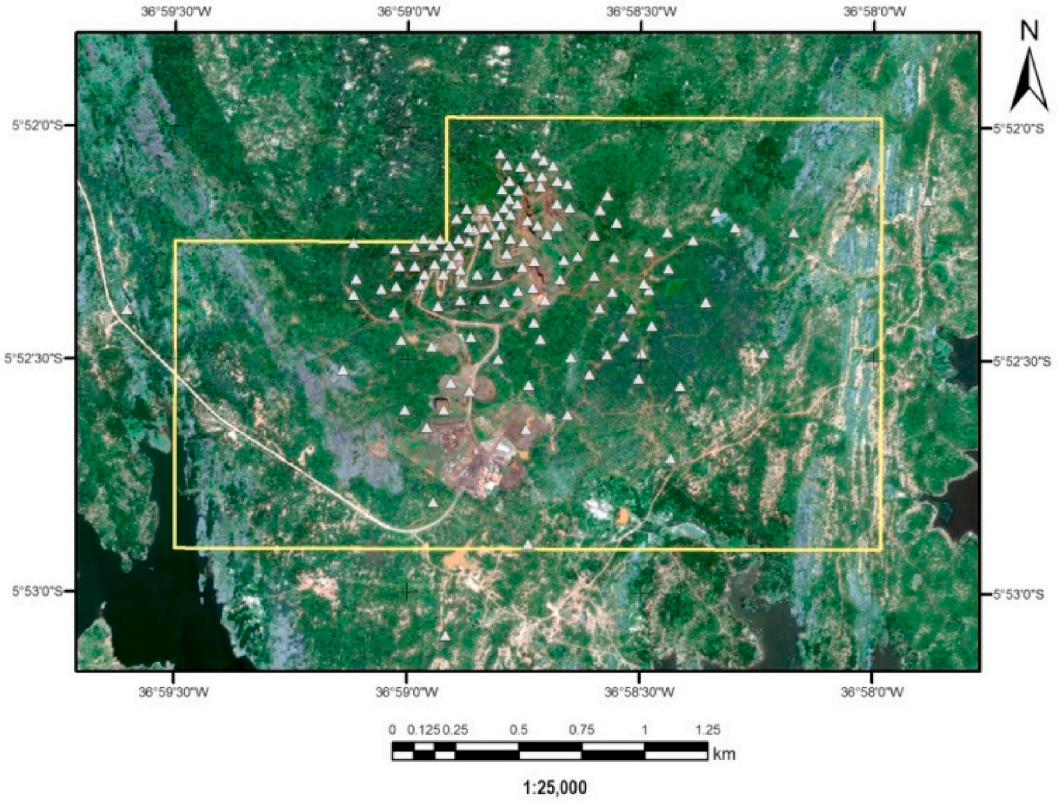

Fig. 6. Map showing the location of the boreholes.

Table 1
Descriptive summary exploratory analysis (%wt).

Variable	Mean	Median	Min.	1st Quant.	3rd Quant.	Max.	Std. Dev.	C.V.
Fe_2O_3	41.68	43.18	4.81	38.01	46.90	70.01	8.54	20.48
SiO_2	56.83	55.68	28.11	52.28	60.17	90.23	7.77	13.67
Al_2O_3	1.134	0.505	0.010	0.250	0.905	18.230	2.26	199.60
P	0.041	0.040	0.007	0.030	0.050	0.252	0.02	55.82
Mn	0.309	0.210	0.010	0.130	0.360	2.030	0.29	95.20

Table 2
– Class limits defined by CA modeling.

Variable	Class	Cum. Freq.	Grade limits%	Average content
Fe_2O_3	Fe1	0.334	4.18–40.60	32.40
	Fe2	0.668	40.60–45.62	43.17
	Fe3	1.000	45.62–70.10	49.50
SiO_2	Si1	0.501	28.11–55.68	51.26
	Si2	1.000	55.68–90.23	62.42
Al_2O_3	Al1	0.517	0.010–0.500	0.27
	Al2	1.000	0.500–18.23	2.06
P	P1	0.547	0.007–0.040	0.03
	P2	1.000	0.040–0.252	0.06
Mn	Mn1	0.368	0.010–0.160	0.10
	Mn2	0.679	0.160–0.300	0.23
	Mn3	1.000	0.300–2.030	0.63

Table 3
- Eigenvalues.

Axis	Eigenvalue	% Explained	% Cumulated
1	0.1845	49.98	49.98
2	0.0604	16.35	66.33
3	0.0472	12.78	79.11
4	0.0357	9.67	88.78
5	0.0234	6.34	95.12
6	0.0167	4.51	99.64
7	0.0013	0.36	100

model comprised 4 clusters while the last model had 12 clusters. Having few variables (in this case, axes 1 and 2), the algorithm was set to perform ten iterations; however, a solution was obtained after 2 iterations.Through K-means clustering method, the BIF samples were segregated into groups around one cluster center onto the factorial plane. Factorial data processing did not require preselected seeds, and the K cluster centers were determined on reaching the final number of iterations. As the input data was displayed in a Cartesian plane, no anomalies were expected within the Euclidean distances.

In section 3.2.2, we briefly introduced some studies concerning validation indexes for the best clustering models. Notwithstanding, following the calculations obtained through the algorithm proposed by Hartigan and Wong, 1979 coupled with the variance ration criterion (VRC) by Calinski and Harabasz, 1974, the K-means clustering outcomes were assessed similarly as in Fisher's one-way analysis of variance (F-test) approach. The maximum values of the VRC index could reveal the most appropriate number of clusters for the typological model. It is noteworthy that the petrographic analysis of the BIF-types helped identify the groups.The clustering tests yielded several results showing that an increasing number of partitions could directly reflect BGSS and

Table 4
- Absolute contributions of the classes. The bold values refer to threshold contributions on the three main axes (AC > 8.3%).

Class	Axis1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7
Fe1	**21.41**	1.87	2.34	2.28	8.83	3.11	26.65
Fe2	0.33	**11.83**	**32.19**	0.09	16.96	5.20	0.01
Fe3	**16.53**	4.31	**17.25**	1.45	1.32	0.27	25.66
Si1	**16.95**	1.59	1.89	2.21	3.41	1.03	22.92
Si2	**17.00**	1.59	1.90	2.21	3.42	1.03	23.00
Al1	0.33	**23.03**	2.10	3.15	12.96	6.33	0.46
Al2	0.35	**24.62**	2.24	3.37	13.85	6.77	0.49
P1	4.28	1.67	2.02	23.73	5.59	8.00	0.02
P2	5.17	2.01	2.44	28.65	6.75	9.66	0.02
Mn1	**9.13**	**11.48**	0.28	6.87	0.25	34.67	0.44
Mn2	0.12	**15.92**	**15.58**	22.53	11.59	3.18	0.02
Mn3	**8.41**	0.09	**19.79**	3.46	15.06	20.76	0.32

WGSS values. Lower and higher values for WGSS and BGSS were estimated, respectively, as K increased. VRC tests were performed for both axes showing the F-statistics estimates. Table 6 shows the main results of K-means clustering.

The K_3, K_4, and K_5 models showed broad sample groupings that were interpreted as unsatisfactory typological models due to an excessive mix of petrographic samples. Additionally, a BIF typological model was expected to appropriately reflect geochemical composition. The K_6 model had the maximum VRC mean among the first eight clustering models. However, cluster No. 6 grouped all HI, MgI1, MgI2, MgI3, MgI5, and MgI7 samples and almost 22% of the total samples. The K_6 model was the first to present a cluster with no associated petrographic sample.The K_7 model had the third lowest VRC mean among the first eight clustering models, although exhibited petrographic samples throughout each cluster. This aspect helps geochemical interpretation; however, it still has lower VRC values. The K_8 model had the third highest VRC mean and presented a cluster with no associated petrographic sample. Moreover, the highest proportion of sample gathering was 17% (clusters No. 4 and No. 6). Models K_9 to K_{12} did not receive any further consideration due to the existence of two clusters (in each model) with no grouped petrographic sample, although their VRC mean were maximized (Table 6). These cluster models make geochemical interpretation considerably challenging and probably helping obtain no information. Considering the above, the K-means model with eight clusters (K_8) provided the most appropriate typological arrangement in terms of balancing geochemical profiles of the samples and the statistical constraints (Table 7 and Fig. 9).

5. Discussion

5.1. CA results

During the CA modeling, a question emerged as to how an appropriate encoding scheme could be obtained by splitting geochemical content of variables into grade limits (Table 2). This is possible by appropriately utilizing the chi-square distance as a measure for the dual variables-samples representation onto a factorial plane. It relies on the fact that variables and samples are *profiles* that are encoded as conditional frequencies (Lebart et al., 1984). According to Benzécri, 1977 and Lebart et al., 1984, an invariance of results is guaranteed by the chi-square metric because of the inverse of the conditional frequency weight for each element of the data matrix. In this matter, situations could arise as the first round of encoding begins, such as in the CA geochemical modeling of the BIF typology. Two classes of the same variable plotted in the same sector were first noticed for SiO_2, Al_2O_3, and P contents (Fig. 7). Lebart et al., 1984 reported an aggregation of similar profiles does not induce a loss of information in certain categories. Meanwhile, increasing the number of factorial divisions of a variable does not improve results.

In terms of the geochemical findings, Figs. 7 and 8 show encoded variables and samples projected onto the CA factorial planes, observations that could be drawn include: 1) Samples rich in Fe_2O_3 were positioned on the positive side of the plane of Axis 1 and they are essentially represented by hematitic and magnetitic itabirites. Samples near Axis 2 are likely to be rich in Al_2O_3. MgI8 was a representative sample of this case; 2) The negative side of Axis 1 represents itabirites depleted in Fe_2O_3. Nevertheless, one magnetitic itabirite (MgI4) was positioned near the amphibolitic itabirites (AmI1, AmI2, and AmI3) with high Al_2O_3 content; 3) Axis 2 defines the amount of Al_2O_3 present in geochemical samples. This variable has an outstanding absolute contribution (Table 4). The locations of samples MgI4, AmI1, AmI2, AmI3, and MI2 is explained by the presence of hornblende (an aluminous amphibole);4) P grades have an insignificant geochemical meaning on plane A1-A2 due to its low contribution. The most significant contributions of this variable are related to Axis 4 alone. Therefore, no valid correlations can be made due to the low percentage of explained variance for that factor (Table 3).

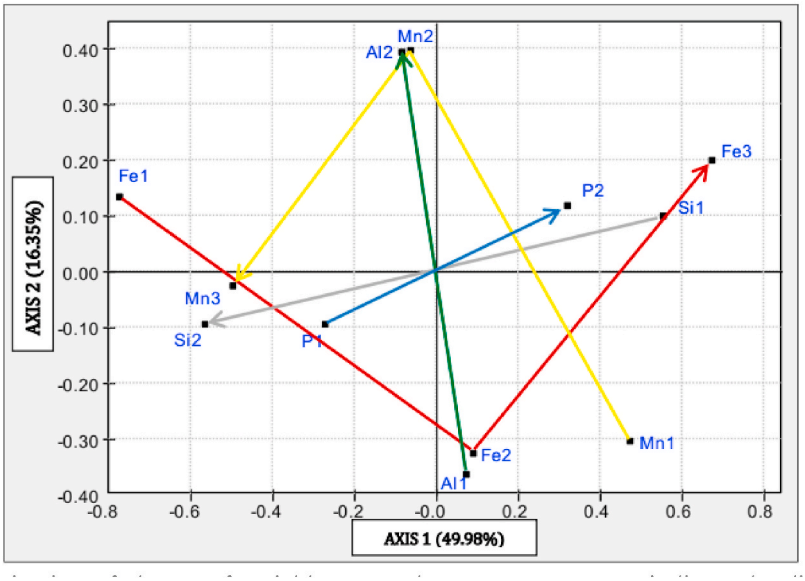

Fig. 7. Projection of classes of variables onto plane A1-A2. Arrows indicate the direction of grade variation according to Table 2.

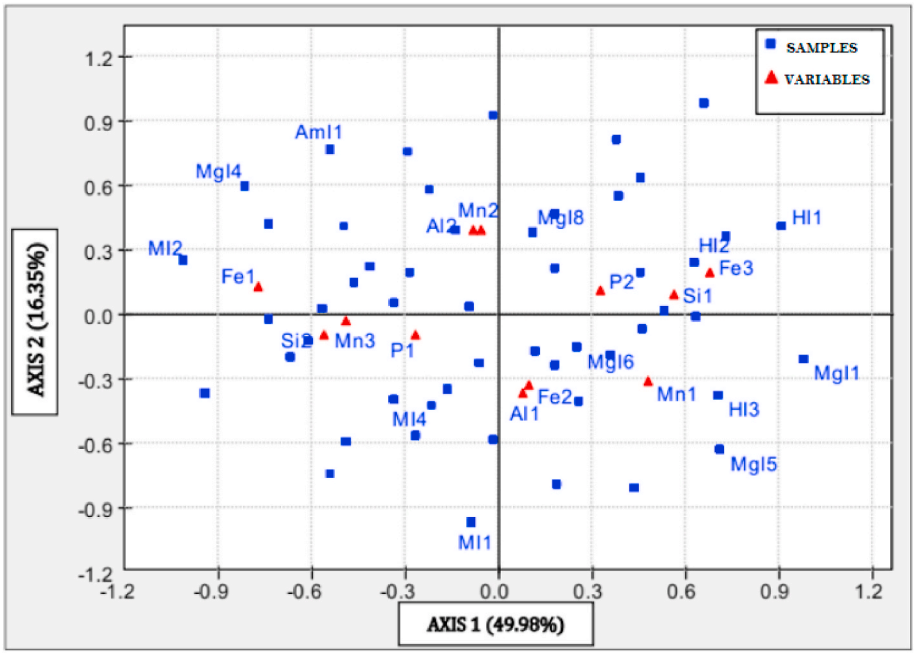

Fig. 8. Projection of classes of variables and samples onto plane A1-A2. Some petrographic samples occupy the same position in the plane: AmI1=AmI2=AmI3, MgI1= MgI2= MgI7, MI1=MI3, and HI3=MgI.

Table 5
Geochemical composition of the petrographic samples.

BIF type	Sample	Fe_2O_3	SiO_2	Al_2O3	P	Mn
Amphibolitic Itabirite	AmI1	18.28	68.96	12.54	0.049	0.17
	AmI2	39.56	57.90	2.25	0.056	0.22
	AmI3	30.75	68.33	0.57	0.057	0.29
Hematitic Itabirite	HI1	52.71	46.38	0.82	0.050	0.04
	HI2	51.83	47.49	0.54	0.030	0.11
	HI3	50.57	49.01	0.26	0.010	0.16
Martitic Itabirite	MI1	43.54	56.12	0.24	0.031	0.06
	MI2	36.78	58.91	3.37	0.021	0.92
	MI3	42.06	57.28	0.49	0.040	0.13
	MI4	42.98	56.72	0.01	0.031	0.26
Magnetitic Itabirite	MgI1	53.98	45.69	0.10	0.072	0.15
	MgI2	46.18	53.52	0.10	0.041	0.16
	MgI3	47.05	52.53	0.34	0.020	0.06
	MgI4	38.21	61.02	0.54	0.032	0.20
	MgI5	45.53	54.16	0.19	0.051	0.07
	MgI6	50.01	48.97	0.17	0.039	0.82
	MgI7	45.96	53.47	0.40	0.052	0.12
	MgI8	45.56	52.61	1.54	0.034	0.25

5.2. Geochemical interpretation

The Mn content in the Bonito mine BIFs is significantly low to form manganese ore minerals. However, the Mn grade could be related to a possible continental input to the basin, through terrigenous sediments deposited due to fluvial systems. Additionally, Mn grades tend to occur

Table 7
Summarized K-means results. The means of clusters were positioned as factorial coordinates on the A1-A2 plane.

Cluster	Cluster means (variable coordinates)		Number of clustered samples	BIF cluster
	Axis1	Axis2		
1	0.07	−0.14	174	MgI6
2	0.23	0.30	113	MgI8
3	−0.16	−0.75	176	MI1-MI3-MI4
4	0.63	0.58	239	HI1–HI2
5	−0.81	0.27	164	MI2
6	0.80	−0.34	239	HI3–MgI1–MgI2–MgI3–MgI5– MgI7
7	−0.65	0.67	105	AmI1-AmI2-AmI3-MgI4
8	−0.86	−0.27	174	X-itabirites

Table 6
VRC results through K_n-means clustering.

n	Axis 1			Axis 2			VRC Mean
	BGSS	WGSS	VRC-A1	BGSS	WGSS	VRC-A2	
3	440.20	154.29	1969.98	168.50	171.56	678.18	1324.08
4	461.23	133.26	1592.10	246.92	93.14	1219.42	1405.76
5	526.15	68.35	2654.01	223.33	116.73	659.58	1656.79
6	539.05	55.44	2679.64	264.64	75.42	967.01	1823.33
7	509.07	85.42	1367.66	295.59	44.47	1525.48	1446.57
8	539.19	55.30	1916.44	292.85	47.22	1219.18	1567.81
9	556.83	37.66	2541.30	305.10	34.96	1499.82	2020.56
10	561.56	32.94	2603.02	306.95	33.11	1415.11	2009.06
11	560.34	34.15	2252.85	311.30	28.77	1485.69	1869.27
12	570.17	24.32	2923.95	315.44	24.63	1597.50	2260.72

n – Number of generated clusters; BGSS – Between groups sum of squares; WGSS – Within-group sum of squares; VRC – variance ratio criterion.

Table 8
Proposed typological classification for the BIFs of the Bonito mine.

Cluster	BIF cluster	BIF typology
1	MgI6	Magnetitic Itabirite (MgI)
2	MgI8	Aluminous-Magnetitic Itabirite (Al–MgI)
3	MI1-MI3-MI4	Martitic Itabirite (MI)
4	HI1–HI2	Hematitic Itabirite (HI)
5	MI2	Aluminous-Martitic Itabirite (Al-MI)
6	HI3–MgI1– MgI2–MgI3–MgI5–MgI7	Hematitic-Magnetitic Itabirite (H–MgI)
7	AmI1-AmI2-AmI3-MgI4	Magnetitic-Amphibolitic Itabirite (Mg-AmI)
8	*X-itabirites*	Silicate-Martitic Itabirite (Si-MI)

along with SiO_2 content and, there is seemingly no relation to Fe_2O_3 enrichment. Oxidation during metamorphism and/or weathering events do not concentrate Mn within rich Fe_2O_3 itabirites. Martitic and amphibolitic itabirites (MI and AmI) could have higher Mn concentrations.

The geochemical behavior of iron and manganese in sedimentary processes are almost independent. As sedimentary processes started forming the BIFs of the Bonito mine, manganese presumably did not precipitate with iron compounds and continued dissolving in water due to specific pH-Eh conditions (Krauskopf, 1957). Grouping magnetitic and hematitic itabirites suggests that oxidation processes could have occurred independently, and hematitic itabirites could have the same metallogenesis based on their geochemical peculiarities. Martitization occurred with magnetitic itabirites as well; this is related to the loss of FeO molecules. Nevertheless, the newly formed iron mineral still retains its magnetic properties. Fig. 10 shows that martitization could be interpreted as a process occurring with an increase in Mn and SiO_2 amounts.

6. Conclusions

- Petrographic studies helped clearly identify four main BIF-types. Throughout the CA/ multivariate data analysis, petrographic information was reconciled with a large geochemical database offering a supervised CA modeling methodology;
- As the main outcome of the BIF typological model proposed here, it can be stated that not all BIF petrographic types were separated into purely distinct groups over a factorial plane. Clearly, CA using geochemical data has shown that some a priori BIF-types are closely associated with each other. Although CA projections clearly demonstrated the distribution of samples throughout the factorial plane along with geochemical classed-variables, a post-processing classification method became imperative for the geochemical groups. Therefore, K-means clustering was considerably suited for the task;
- The choice of Calinski and Harabasz's criterion was justified by the results obtained by others (Milligan and Cooper, 1985; Saitta et al., 2008; Liu et al., 2010) when comparing some of the best-known validity indexes. However, analyzing VRC values brought up some issues usually not addressed in specific K-means clustering studies; one that became apparent was interpreting the VRC or any validity index for uncorrelated orthogonal axes as variables. In our approach, choosing the K_8-means partition as a BIF typological model was aided by petrographic analyses;
- It was impossible to verify whether martitization occurred due to medium-high temperatures as in amphibolite facies metamorphism

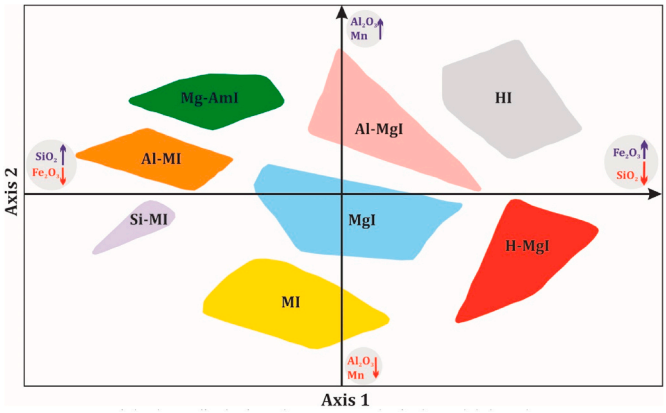

Fig. 10. Factorial plane displaying the BIF typological model based on K-means post-processing. Gray-circles with geochemical variables depict the relative enrichment (upward blue-arrow) or depletion (downward red-arrow) of chemical content. Phosphorus content was unrepresented due to its low statistical significance in this factorial setting. BIF typology acronyms can be referred to from Table 8.

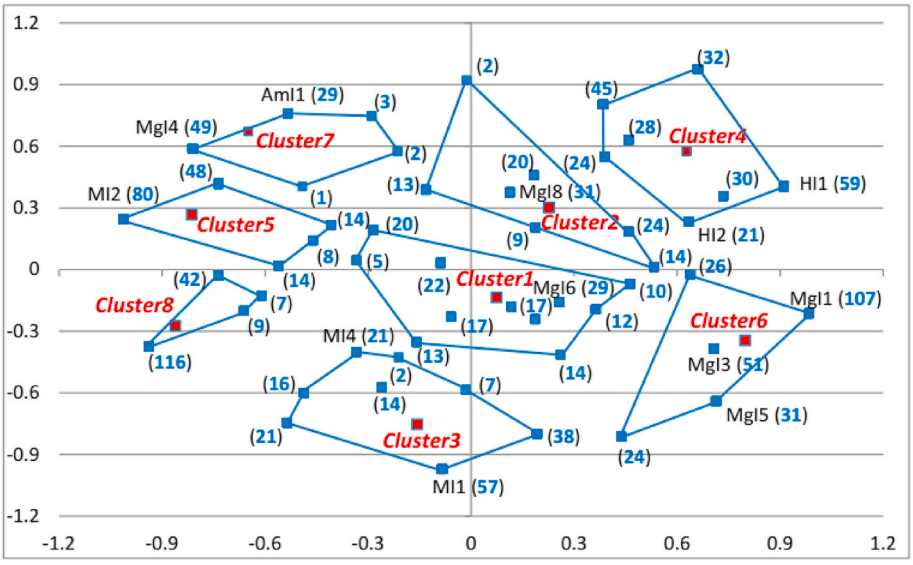

Fig. 9. Plane A1-A2 showing the K-clusters for the Bonito Mine BIF typology. Numbers between parentheses indicate the number of samples grouped by each point in the plane. Red squares depict the centroid of each delimited cluster.

or due to strong weathering, as has been reported for other BIF-ore deposits (Barros, 1963; Ramdohr, 1981; Veríssimo, 1999; Klein, 2005). Martitic itabirites tended to be dominant in the negative sector of Axis 1, indicative of an increase of silicates on that sector of the factorial plane.

7. Future work

The BIFs of the Serra dos Quintos Formation are now advantageously understood in terms of geological processes. Typology modeling could help disclose aspects useful for enhancing geological modeling, mining planning, ore processing, etc.

Further research could consider the spatial distribution of BIF-types throughout the study area. Once the BIF typological model is established and geochemical samples are typologically known, such a novel geological data ensemble could be available for 3D typological modeling. Thus, the entire itabiritic body could be modeled and volumetric information could be evaluated.

To a better understand ore-forming processes, iron isotope studies on itabirites focusing on chemical fractioning considering the initial depositional processes until major metamorphic events could be conducted.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Brasil (Grant# 8887.116689/2016–00). The first and second authors are greatly indebted to Mr. Moacir Dantas, Chief-engineer for MHAG Serviços e Mineração S.A. company who allowed the access to the Bonito mine facilities and giving us the kind permission for using their database. Special acknowledgements are due to Prof. Dr. Manuel F. C. Pereira, Chief of the Geosciences Museums of Instituto Superior Técnico at the Universidade de Lisboa/Portugal, for cooperation and providing the conditions for the petrographic studies and to Prof. Dr. António Jorge G. de Sousa of CERENA/IST for crucial discussions and suggestions. We thank Mr. Joel Pedrosa at the Electronic Microscopy Laboratory of the Departamento de Geologia, Universidade Federal do Ceará for the elaboration of the petrographic thin/polished sections. We also thank the anonymous reviewers for their suggestions and helpful comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeochem.2020.104779.

References

Amorim, R.C., 2016. Survey on feature weighting based K-means algorithms. J. Classif. 33 (2), 210–242.
Angelim, L.A.A., Medeiros, V.C., Nesi, J.R., 2006. Geologia e recursos minerais do estado do Rio Grande do Norte: texto explicativo dos mapas geológico e de recursos minerais do estado do Rio Grande do Norte. Recife: CPRM; FAPERN, 2006. Escala 1: 500.000. Programa Geologia do Brasil (PGB)., , 1st1. CPRM-SGB, Recife.
Archanjo, C.J., Viegas, L.G.F., Hollanda, M.H.B.M., Souza, L.C., Liu, D., 2013. Timing of the HT/LP transpression in the neoproterozoic Seridó Belt (borborema province, Brazil): Constraints from U-Pb (SHRIMP) geochronology and implications for the connections between NE Brazil and West Africa. Gondwana Res. 23, 701–714.
Balladur, J.P., 1970. Analyse factorielle des correspondences. Annales de l'Inséé 4, 47–79.
Barbosa, I.G., 2013. Mina do Bonito – Tipologia e geoquímica dos minérios de ferro, Jucurutu/RN – Brasil. M.Sc. Dissertation. Department of Geology, Federal University of Ceará. Fortaleza.
Barros, L.A., 1963. Mineralogia dos jazigos de ferro e manganês de Goa. Unpublished Full Professor Thesis. Superior Technical Institute, University of Lisbon, Lisbon.
Benzécri, J.P., 1973. L'Analyse des Données, 2. L'Analyse des Correspondences. Dunod, Paris.
Benzécri, J.P., 1977. Histoire et préhistoire de L'Analyse des Données - partie V: L'analyse des correspondences. Cah. Anal. Données 2 (1), 9–40.
Bezdek, J.C., Pal, N.R., 1998. Some new indexes of cluster validity. IEEE Trans Sys Man Cyber 28 (3), 301–315.
Birks, H.J.B., 1987. Multivariate analysis in geology and geochemistry: An introduction. Chemometr. Intell. Lab. Syst. 2, 15–28.
Calinski, R.B., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. 3, 1–27.
Craig, J.R., Vaughan, D.J., 1984. Ore Microscopy and Ore Petrography, second ed. John Wiley and Sons, Inc., New York.
CVRM, 2002. Programa ANDAD 7.20 - manual do usuário. Centro de Geosistemas. IST. Lisbon. Website for download (software and manual): http://biomonitor.ist.utl.pt/~ajsousa/Andad.html. (Accessed 13 June 2019).
Ferreira, C.A., Santos, E.J., 2000. Programa Levantamentos Geológicos Básicos do Brasil. Jaguaribe SE. Folha SB. 24-Z. Estados do Ceará, Rio Grande do Norte e Pernambuco. Scale 1: 500.000. Geological Survey of Brazil - CPRM 1 (1), 1–50. Recife. (CD ROM).
Grenacre, Michel, Blasius, Jörg, 2006. Multiple Correspondence Analysis and Related Methods: Methods. Statistics in the Social and Behavioral Sciences Series. Chapman & Hall/CRC, Boca Raton.
Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-means clustering algorithm. Appl Statist 28 (1), 100–108.
Hill, M.O., 1974. Correspondence analysis: a neglected multivariate method. Appl Statist 23 (3), 340–354.
Hollanda, M.H.B.M., Archanjo, C.J., Bautista, J.R., Souza, L.C., 2015. Detrital zircon ages and Nd isotope compositions of the Seridó and Lavras da Mangabeira basins (Borborema Province, NE Brazil): evidence for exhumation and recycling associated with a major shift in sedimentary provenance. Precambrian Res. 258, 186–207.
Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. 31, 651–666.
Klein, C., 2005. Some Precambrian banded iron-formations (BIFs) from around the world: their age, geologic setting, mineralogy, metamorphism, geochemistry, and origin. American Mineralogist 90, 1473–1499.
Krauskopf, K.B., 1957. Separation of manganese from iron in sedimentary processes. Geochem. Cosmochim. Acta 12, 61–84.
Lebart, L., Morineau, A., Warwick, K.M., 1984. Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Methods for Large Matrices. John Wiley and Sons, New York.
Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., 2010. Understanding of internal clustering validation measures. In: IEEE Int Conference on Data Mining, Sydney, NSW, pp. 911–916.
MHAG Serviços e Mineração, S.A., 2013. Projeto Produção de Pellet Feed 2.0 MPTA: Mapa geológico de detalhe. Unpublished Internal Report, 3rd Revision. Jucurutu, Brazil.
Mellinger, M., 1987. Correspondence Analysis: the method and its application. Chemometr. Intell. Lab. Syst. 2, 61–77.
Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika 50 (2), 159–179.
Parks, J., 1966. Cluster analysis applied to multivariate geological problems. J. Geol. 74, 703–715.
Patinha, C., Correia, E., Silva, E.F., Simões, A., Reis, P., Morgado, F., Fonseca, E.C., 2007. Definition of geochemical patterns on the soil of Paul de Arzila using correspondence analysis. J. Geochem. Explor. 98 (1/2), 34–42.
Pereira, H.J.F.G., 1981. Análise estrutural e seus reflexos na avaliação económica de recursos minerais. Ph.D. Thesis. Superior Technical Institute, University of Lisbon, Lisbon.
Pereira, H.G., Sousa, A.J., Ribeiro, J.T., Salgueiro, R., Dowd, P., 2015. Correspondence Analysis as a Modeling Tool. IST Press, Lisbon.
Ramdohr, P., 1981. The Ore Minerals and Their Intergrowths, second ed. Pergamon Press, Oxford.
Reis, A., Sousa, A.J., Silva, E., Patinha, C., Fonseca, E.C., 2004. Combining multiple correspondence analysis with factorial kriging analysis for geochemical mapping of the gold-silver deposit at Marrancos (Portugal). Appl. Geochem. 19 (4), 623–631.
Rhodes, J., 1969. The application of cluster and discriminatory analysis in mapping granite intrusions. Lithos 2, 223–237.
Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to interpretation and validation of cluster analysis. Jour Comp Appl Math 20, 53–65.
Sá, E.F.J., Salim, J., 1980. Reavaliação dos conceitos estratigráficos na região do Seridó (RN-PB). Miner Metal 80, 16–28.
Sá, E.F.J., 1994. Faixa Seridó (Província Borborema, NE do Brasil) e o seu significado geodinâmico na Cadeia Brasiliana/Pan-Africana. Ph.D. Thesis. Institute of Geosciences, University of Brasília, Brasília.
Sá, E.F.J., Fuck, R.A., Macedo, M.H.F., Peucat, J.J., Kawashita, K., Souza, Z.S., Bertrand, J.M., 1995. Pre-Brasiliano orogenic evolution in the Seridó Belt, NE Brazil: Conflicting geochronological and structural data. Rev. Bras. Geociencias 25 (4), 304–314.
Saitta, S., Raphael, B., Smith, I.F.C., 2008. A comprehensive validity index for clustering. Intell. Data Anal. 12, 529–548.
Sial, A.N., Campos, M.S., Gaucher, C., Frei, R., Ferreira, V.P., Nascimento, R.C., Pimentel, M.M., Pereira, N.S., Rodler, A., 2015. Algoma-type Neoproterozoic BIFs and related marbles in the Seridó Belt (NE Brazil): REE, C, O, Cr and Sr isotope evidence. J. S. Am. Earth Sci. 61, 33–52.
Statsoft Inc, 2010. Statistica 10 Enterprise User's Manual. Tulsa.

Teil, H., 1975. Correspondence factor analysis: an outline of its method. Math. Geol. 7 (1), 3–12.
Teil, H., Cheminee, J.L., 1975. Application of correspondence factor analysis to the study of major and trace elements in the Erta Ale Chain (Afar, Ethiopia). Math. Geol. 7 (1), 13–30.
Teillard, P., Volle, M., 1976. Détection des points aberrants en analyse factorielle des correspondances. Annales de l'Inséé 22/23, 237–254.
Valenchon, F., 1982. The use of correspondence analysis in geochemistry. Math. Geol. 14 (4), 332–342.
Van Schmus, W.R., Neves, B.B.B., Williams, I.S., Hackspacher, P.C., Fetter, A.H., Dantas, E.L., Barbinski, M., 2003. The Seridó Group of NE Brazil, a late Neoproterozoic pre- to syn-collisional basin in West Gondwana: insights from SHRIMP U-Pb detrital zircon ages and Sm-Nd crustal residence (T_{DM}) ages. Precambrian Res. 127 (4), 287–327.
Veríssimo, C.U.V., 1999. Jazida de Alegria: Gênese e tipologia dos minérios de ferro (Minas 3,4 e 5 - Porção Ocidental). Ph.D. Thesis. Institute of Geosciences and Mathematical Sciences, Júlio de Mesquita State University of São Paulo, Rio Claro.