Hyperspectral Imaging, Data Processing and Prediction of Bauxite Samples from Paragominas, Brazil

Acácio N. Pina Neto[1], Éricka Brito[2], Gustavo Loureiro[3], Ricardo Radtke[4], Bruno Gomes[5] and Silvia Farias[6]
1, 3, 4. Exploration Geologists
2. Database Analyst
5. Mineral Exploration Manager
6. Mineral Exploration Senior Manager
Hydro Bauxite & Alumina, Paragominas, Brazil
Corresponding author: acacio.pina.neto@hydro.com
https://doi.org/10.71659/icsoba2024-bx002

Abstract

The use of hyperspectral sensors has become increasingly popular for mapping ground targets. Due to the large number of spectral bands present in these sensors and the narrow spectral bandwidth, mineralogical identification and quantification are possible with precision according to the spectral signature of each mineral through a known and specific pattern of reflection and absorption at different wavelengths. To evaluate the potential use of these sensors in bauxite mining areas, mining fronts were imaged in Paragominas, in the northeast region of the State of Pará, Brazil, using the HyspeX hyperspectral sensor (SWIR 970–2500 nm). 72 samples were collected for scanning and construction of the spectral library with the assistance of a FieldSpec 3 Jr. spectroradiometer that captures 2151 bands between 350 and 2500 nm (vis-NIR). The chemical results were subjected to statistical and discriminant analysis using SAS software, and the images were processed in CaliGeo PRO and ENVI Classic software. Algorithms developed in the Python language (libraries available in Scikit Learn, MatLab, and Orange software) were used to evaluate, model, and predict the spectral curves with the chemical contents of the samples. The results showed a moderate distinction between the layers evaluated by the reflectance spectrum in the laboratory, both in shape and intensity. The use of neural networks for prediction presented the best results (60 to 90 % accuracy), and the spectral range of 1000–2500 nm established more robust models for the analyzed data, attesting to the potential for application of the hyperspectral methodology for characterization and quality control of bauxite mining. However, a larger set of samples and a more robust spectral library would allow refinement and improvement of the prediction model's performance, reducing the areas of overlap in the spectral responses.

Keywords: Hyperspectral imaging, Model prediction, Bauxite exploration.

1. Introduction

The use of hyperspectral sensors has become increasingly popular for mapping ground targets. Due to the large number of spectral bands present in these sensors and their narrow spectral bandwidth, they can accurately provide information that is not obtained with multispectral sensors. Different minerals have a specific pattern of reflection and absorption at a wide range of wavelengths. In this way, many minerals can be identified according to their spectral response, which in some cases is specific to each species, also known as spectral signature [1].

Over the past three decades, there have been great advances in imaging hyperspectral data with improved identification and quantification capacity for geological materials on the earth's surface [2, 3, 4]. However, most of the works where these images are used refer to those obtained by sensors installed on space platforms, such as the Hyperion sensor, and more recently, data acquired from platforms installed on airplanes.

Through the years, imaging technological advances have allowed the acquisition of data from sensors mounted on platforms based in the field [5,6,7]. This condition made it possible to use images in open-pit mining areas since the equipment has become increasingly smaller and easily transportable. The use of these imaging sensors has great potential for applicability because it allows the identification and, in some cases, the quantification or abundance of minerals in mining sites [8, 9]. This starts to play a relevant role when, increasingly, we seek, within the exploration process, methods that can reduce the exposure of people who carry out inspection, collection, and evaluation of material within mined areas, whether for reasons of safety or health [10]. Furthermore, the increased use of images obtained by remote sensors installed on platforms that can be operated in the field has been used, particularly, for the possibility of evaluating and obtaining values in a non-destructive way in areas that require minimum spatial scale.

In this context, this project aimed to evaluate the potential use of hyperspectral imaging sensors in bauxite mining owned by Hydro Paragominas, in the northern region of Brazil, and present the positive and negative points for possible automation in the process assessment of the quality of the mined material.

2. Methodology

2.1 Study Area

The study area is located in the central domain of the Paragominas Bauxite Province (PBP), in the northeast region of the state of Pará, in the Eastern Amazon, Brazil. The PBP represents one of the most important, extensive, and dense groupings of bauxite deposits in Brazil, with a potential of more than 3 billion tonnes of metallurgical ore, about 70 % of Brazil's total bauxite reserves. The PBP is characterized by plateaus covered by a thick layer of clay (Belterra clay) and ferroaluminous crusts. These deposits were formed by the lateritic alteration of siliciclastic deposits from the Cretaceous, specifically sediments from the Itapecuru and the Ipixuna Formation, during the Paleogene [11] (see Figure 1).

The central domain of the Paragominas Province comprises several bauxite deposits, including the Miltonia 5 and Miltonia 3 plateaus, where the Hydro Paragominas bauxite mine has been in operation since 2012. The mine is located in the municipality of Paragominas, 356 km from the capital, Belém do Pará. The lateritic profile consists of eight main lithotypes with clear textural, compositional, and color differences and well-defined contacts. Starting from the bottom to the top of the lateritic profile, the lithotypes are as follows: a) The bottom clay (ARV), which transitions gradually from the fine-grained kaolinitic sandstone bedrock of the Itapecuru/Ipixuna formations. This layer includes a saprolitic and mottled zone ending in the main bauxite zone; b) The amorphous bauxite (BA) is found in pseudo nodules and columnar features with a microcrystalline texture, along with a mottled clay matrix. It represents the transition from the bottom clay to the main ore zone; c) The ore zone is a massive layer of reddish bauxite containing numerous gibbsite crystals and iron oxides. It is divided into crystallized bauxite with "amorphous" bauxite (BCBA) and crystallized bauxite (BC); d) Ferruginous Laterite (LF) is composed of goethite and hematite pisolites in a massive texture; e) Crystallized Nodular Bauxite (BNC) presents gradual contact with the horizon above and discordantly overlaps with the level of ferruginous laterite. It is characterized by concretions of irregular size and shape, composed of crystallized gibbsite and iron oxides, and commonly occurs as lenticular and discordant bodies; f) Nodular Bauxite (BN) consists of gibbsite nodules formed by amorphous bauxite in a kaolinitic matrix; g) Belterra clay (CAP) is discordant and contains ferruginous pisolites among gibbsite nodules. This geological formation covers the lateritic profile and is 5 to 20 meters thick, consisting of a homogeneous sequence of kaolinitic clays (see Figure 2).

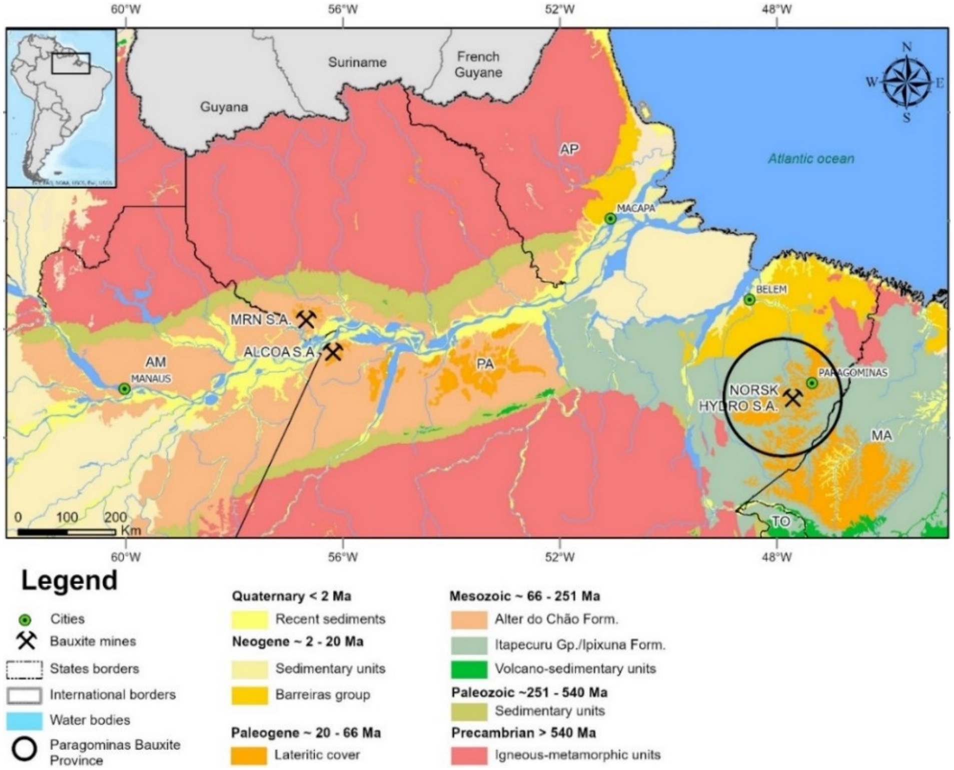


Figure 1. Simplified geological map of northern Brazil showing the distribution of the main lithostratigraphic units.

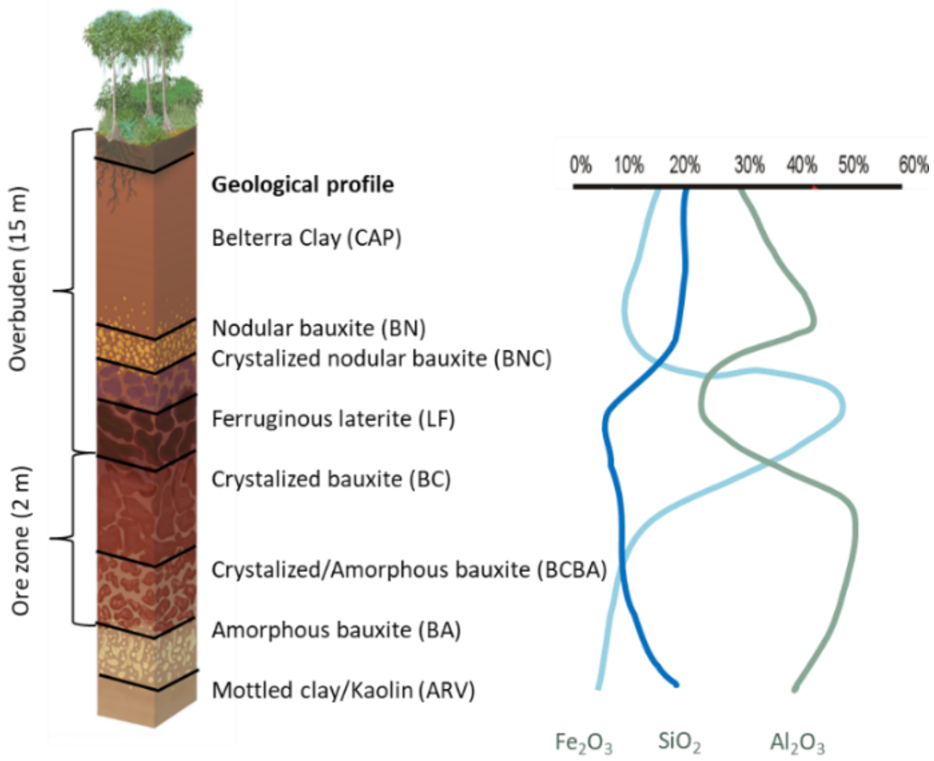


Figure 2. Regional geological profile.

2.2 Hyperspectral Imaging System

The HyspeX hyperspectral imaging sensor was used for slopes scanning in bauxite mining. The system includes a hyperspectral imaging sensor with a detector integrated into a Mercury-Cadmium Tellurite (MCT) optical system, along with a liquid helium cooling system. It captures information in the range of 970–2500 nm with a spectral resolution of 5.1 nm, enabling imaging with 300 spectral bands in the SWIR (short-wave infrared wavelength). The system was mounted on a tripod for field surveys, with a rotating system controlled by the operator and an automated mechanism for direction control (pan) and manual control for lateral variation (tilt) (see Figure 3). A 14-volt battery system powers the imaging unit with a consumption of 1 ampere-hour (Ah). Data from the imaging unit is transmitted to a display unit. By integrating the sensor and the optical system on a single platform, the system allows for continuous imaging without the need for radiometric or geometric calibration.

Figure 3. HyspeX sensor system and its components.

2.3 Field Survey Materials and Methodology

To collect and record samples of mined material, augers, plastic bags, labels, digital cameras, and a wooden board covered with barium sulfate as a white reference for converting image data into reflectance were used. For imaging the slope, it was critical to have an open sky, free of clouds, for good lighting. Imaging was consistently started at 9:35 a.m. and continued until 10:00 a.m. daily. Three sampling points were marked where samples were collected from the slope. To establish a relationship between the collected points and the location from which the reflectance values of the images were extracted, the locations for collecting mineral material were demarcated with ecological paint (see Figure 4).

Figure 4. Imaged slope showing standard barium sulfate plates and locations (X markings) from where samples were taken for laboratory analysis of mineral content.

After finding an appropriate lighting condition, imaging was conducted to obtain data for each scan as specified. Subsequently, samples were gathered from pre-designated points using labeled bags to identify each collected sample (see Figure 5). The collected material was then sent to the laboratory for analysis. The results, along with laboratory imaging and spectroradiometric data, were collected to create the spectral library.

Figure 5. Sampling at pre-defined points for laboratory analysis.

2.4 Spectral Reading in the Laboratory

To obtain the spectra and build a spectral library, readings were conducted using a spectroradiometer and hyperspectral cameras from the Geoprocessing and Remote Sensing Laboratory (GALeS) at the State University of Maringá (UEM). The bauxite samples collected were sieved using a 2 mm sieve and placed in Petri dishes measuring 9 cm in diameter and 1.5 cm in height. Each sample was scanned using a FieldSpec 3 Jr. spectroradiometer that collects in the spectral range of 350–2 500 nm, totaling 2 151 bands, and programmed to perform 30 readings per sample, generating the average spectral curve of each sample analyzed. The geometry chosen for the spectral reading followed the recommendations of Nanni *et al.* [12] (Figure 6).

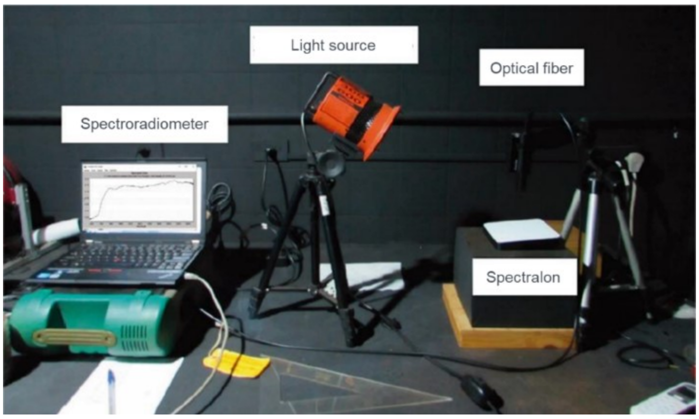


Figure 6. Use of FieldSpec 3 Jr in the laboratory to acquire spectral data from bauxite samples.

The samples were positioned one by one, 8 cm below the device's fiber optic sensor, taking an area of approximately 2 cm^2 as a reading of the soil. For lighting, a 650 W lamp was used, with a beam not collimated to the target plane, positioned 35 cm from the platform and at an angle of 30° with the horizontal plane. Before starting data collection in the laboratory, the device was calibrated and optimized using a standard white plate with 100 % reflectance (Spectralon). Three sequential readings were performed on each sample with a 120° rotation between each reading, and the average was calculated according to Nanni and Demattê [13]. The collected spectra were subjected to correction using the "splice correction" procedure to eliminate steps (or offsets) from the original spectra in the spectral bands of the spectroradiometer operation: VisNIR (350–1 100 nm), SWIR (1 100–2 500 nm), and SWIR 2 (1 800–2 500 nm).

For a better understanding of the methodological procedure, it is presented in the flowchart in Figure 7.

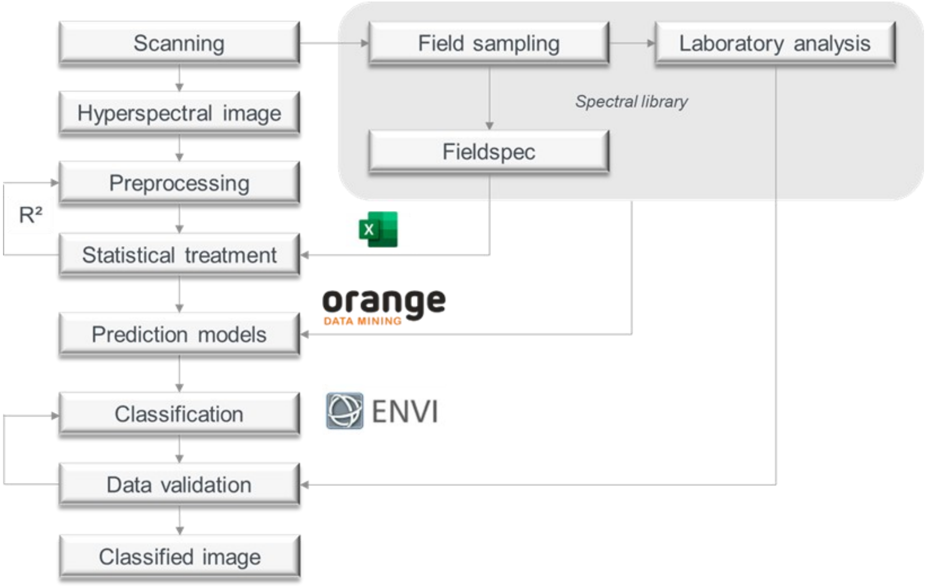


Figure 7. Methodological procedure flowchart.

2.5 Statistical Analysis

2.5.1 Exploratory Data Analysis

The chemical analysis data included measurements of Available Alumina (Al_2O_3 Av.), Reactive Silica (SiO_2 reat.), Total Alumina (Al_2O_3), Iron oxide (Fe_2O_3), Silicon dioxide (SiO_2), Titanium dioxide (TiO_2), as well as Mass Recovery (MR) determined after wet sieving with a 20# (approximately 0.841 mm) opening grid. The sample results were analyzed descriptively, utilizing histograms to assess the distribution of the data. Normality was tested using the Shapiro-Wilk test, and homogeneity of variances was assessed using the Bartlett test at the 5 % significance level with the assistance of the XLSTAT package in Excel software. Prior to commencing the modeling process, the spectral data underwent principal component analysis (PCA) to examine the dispersion of reflectance values of the samples and to distinguish differences between spectra of samples gathered from various lithological conditions and depths. This analysis aimed to identify sets of spectra that were homogeneous, as well as to detect any outliers that could potentially impact the development of prediction models [14].

2.5.2 Discriminant Analysis

Once the data sets were subjected to normality and homoscedasticity tests related to error variance, and the assumptions were met, the data matrices x (spectral curves obtained by the sensors) and y (attributes of the bauxite samples) were submitted for discriminant analysis in SAS software [15]. The discriminant functions were obtained through the "proc STEPDISC" procedure at a 5 % probability level, allowing the selection of the variables that best explain the differences between the analyzed lithologies by selecting the best wavelengths from the entire reflectance spectrum [16]. Once the wavelengths that best explain the differences between the plots and treatments studied were obtained, discriminant analysis was carried out using the "proc DISCRIM" command based on these wavelengths.

This discriminant analysis method allows the creation of a predictive model for the association of areas with different coverage and between depths within the same treatments based on the spectral characteristics observed in each case. The procedure generated discriminating functions for each lithology studied through linear combinations of the predictor variables [17]. The number of discriminant functions can provide a statistically significant degree of variation between lithologies, essentially defining the dimensionality of the discriminant space. An error and success relationship between the evaluated lithologies, resulting from the present attributes, was plotted in a correlation matrix (cross-validation) to evaluate the errors and successes between samples. In addition to the discriminant analysis, the analysis of variance (ANOVA) was carried out using the XLSTAT package in Excel software to define the similarities between the lithologies resulting from the quantities of minerals present.

2.5.3 Bauxite Samples Prediction Models by Regression

To assess the quality of modeling and predictive capabilities of spectral curves related to the minerals present in the samples, multivariate regression methods were utilized to develop chemometric models. The most prominent methods currently used to evaluate sample attributes include PLS, Random Forest, Gradient Boost, Linear Regression Model (LRM), Principal Component Regression (PCR), and Support Vector Machine (SVM) [18]. These models were processed using algorithms developed in Python language and libraries available in the Orange software (Figure 8). Algorithms typically function by extracting information from the spectral matrix dataset (wavelengths) and correlating it with information from the reference dataset (attributes). The mathematical calibration model must be optimized through a cross-validation process where a sample is taken from a series of n samples used in calibration to predict the concentration of the removed sample. This process is repeated until all samples have been excluded once, resulting in a model with the lowest cross-validation error ($RMSE_{CV}$) [19].

During the calibration phase, modeling utilizes information from both the data matrix X (spectral data) and the concentration matrix Y (analytical data). Each algorithm operates differently, but the goal is to obtain predictive models with the lowest root mean square error (RMSE), highest R^2 values, and minimal trend deviation (bias). After each data processing step, the values obtained from laboratory analysis were compared with the model's values to interpret and evaluate the potential of using spectral data as predictors of the variables of interest. To maximize the predictive capacity of the spectral curves, the sample calibration and validation process was conducted using sets of spectral curves (vis/NIR-SWIR) tested separately and together.

2.6 Field and Laboratory Image Classification

The images collected in the field and at the laboratory were processed to classify the lithologies of interest. A set of regions of interest for each image was generated in the ENVI software, and the mean, standard deviation, maximum, and minimum values were calculated for each set.

After implementing a set of representative curves for each lithology (field images) and sample (laboratory), the classification process was carried out using various algorithms. Supervised classification algorithms such as Euclidean distance, neural networks, spectral angle, and support vector machine were utilized.

Finally, the images were saved in TIFF format, along with the average curves for each proportion based on the selected criteria.

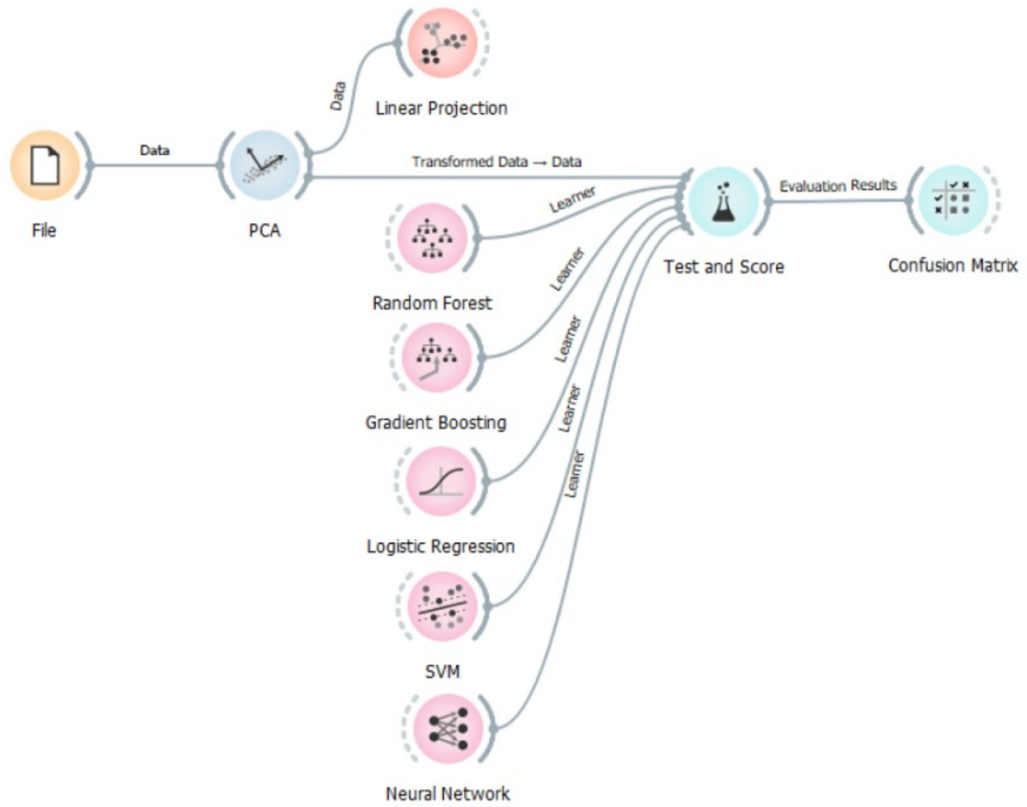


Figure 8. Data preparation and processing flowchart for generating models for predicting sample attributes and their spectral response.

3. Results and Discussion

In total, two areas were used for evaluation. Scenes with lighting problems and the presence of shadows were discarded. 72 samples from different mining points were sent to GALeS-UEM, where they were duly identified for generating prediction and discrimination models.

The results of the sample analysis are presented in Table 1. Descriptive statistical analysis was carried out using the data, evaluating the mean, median, standard deviation, coefficient of variation (CV) and variance. Relatively low dispersion in the data sets is noted for a slight majority of attributes evaluated, except for Al_2O_3 Av., Al_2O_3 and Fe_2O_3. The evaluated conditions presuppose homogeneity of samples, which, in the first analysis, may restrict the quality of the generated models since their quantitative variation was low.

Al_2O_3 av., Al_2O_3 and Fe_2O_3 showed higher dispersion in BN and LF layers. This can be attributed to the transitional gradation between these two layers, as well as contamination by iron content. Iron and alumina typically exhibit a negative correlation.

The attribute data were also subjected to the basic assumptions of normality (Table 2) and analysis of variance (ANOVA) by applying Levene's tests complemented by the mean test between lithologies (Table 3).

Table 1. Descriptive statistical analysis of the chemistry and processing test results of samples collected in the field at different sampling points.

Statistics	Lithology	Nbr. samples	Mean	Std. Dev.	Variance	Coefficient of Variation	Minimum	1st quartile	Median	3rd quartile	Maximum
Al_2O_3 Av.	CAP	5	34.50	1.79	3.22	5.20	31.49	32.96	34.87	35.85	35.91
	BN	14	22.52	13.99	195.77	62.13	2.87	9.35	21.63	33.44	46.89
	BNC	1	50.84	-	-	-	50.84	-	50.84	-	50.84
	LF	16	17.62	8.90	79.17	50.51	4.63	10.16	17.52	25.64	31.66
	BC	14	52.59	2.37	5.61	4.51	47.37	51.57	52.88	54.06	56.11
	BCBA	13	53.14	3.12	9.72	5.87	46.13	50.87	53.31	55.79	56.73
	BA	9	46.57	3.00	8.98	6.43	41.71	43.77	46.93	48.81	51.08
SiO_2 Reat.	CAP	5	12.72	0.76	0.58	5.97	11.75	12.08	12.67	13.38	13.83
	BN	14	10.53	2.91	8.45	27.61	6.55	8.39	10.04	12.76	16.16
	BNC	1	5.73	-	-	-	5.73	-	5.73	-	5.73
	LF	16	6.29	3.19	10.17	50.73	0.87	4.30	5.47	8.53	12.46
	BC	14	2.20	1.01	1.02	45.94	0.76	1.05	2.22	3.19	3.69
	BCBA	13	3.82	1.33	1.77	34.83	1.75	2.63	4.23	4.88	5.64
	BA	9	9.09	2.09	4.39	23.03	4.73	7.95	9.57	10.53	12.01
Al_2O_3	CAP	5	49.19	2.28	5.22	4.64	45.58	47.35	49.67	50.79	51.90
	BN	14	34.12	12.01	144.17	35.19	17.71	24.12	32.85	43.72	55.05
	BNC	1	57.84	-	-	-	57.84	-	57.84	-	57.84
	LF	16	24.45	5.49	30.19	22.47	17.41	19.40	22.93	29.18	34.61
	BC	14	54.70	3.10	9.63	5.67	48.88	52.86	54.41	56.45	59.91
	BCBA	13	57.55	2.89	8.34	5.02	49.96	56.41	58.27	59.50	60.66
	BA	9	56.99	2.15	4.63	3.78	52.91	55.18	58.14	58.77	58.97
SiO_2	CAP	5	14.34	0.86	0.75	6.02	13.32	13.43	14.60	15.13	15.13
	BN	14	12.13	3.61	13.03	29.75	6.90	9.79	11.24	16.30	18.50
	BNC	1	5.84	-	-	-	5.84	-	5.84	-	5.84
	LF	16	7.47	3.92	15.38	52.52	1.41	4.86	6.01	11.12	15.38
	BC	14	2.62	1.09	1.19	41.68	1.09	1.39	2.75	3.66	4.43
	BCBA	13	4.54	1.51	2.29	33.35	2.18	3.35	4.64	5.62	7.56
	BA	9	10.31	2.07	4.28	20.05	6.70	9.35	10.05	11.44	14.36
Fe_2O_3	CAP	5	9.6	3.6	12.6	37.0	6.9	7.5	8.0	12.5	15.8
	BN	14	34.8	15.8	248.1	45.3	8.4	21.6	38.0	49.7	58.4
	BNC	1	6.6	-	-	-	6.6	-	6.6	-	6.6
	LF	16	53.0	8.2	67.1	15.5	36.3	47.4	53.9	58.5	65.7
	BC	14	10.8	4.2	17.3	38.6	4.3	7.9	11.2	13.4	17.9
	BCBA	13	6.2	3.7	13.5	59.5	2.4	3.2	5.4	7.9	16.2
	BA	9	3.8	2.4	5.6	62.4	2.2	2.5	2.8	4.3	9.8
TiO_2	CAP	5	1.61	0.07	0.01	4.42	1.54	1.54	1.59	1.68	1.68
	BN	14	1.16	0.14	0.02	11.70	0.92	1.11	1.16	1.22	1.43
	BNC	1	1.40	-	-	-	1.40	-	1.40	-	1.40
	LF	16	0.95	0.22	0.05	23.21	0.70	0.76	0.97	1.09	1.54
	BC	14	1.85	0.66	0.43	35.51	1.14	1.28	1.71	2.38	3.15
	BCBA	13	1.41	0.48	0.23	33.75	0.99	1.08	1.14	1.85	2.27
	BA	9	1.04	0.37	0.14	35.84	0.71	0.83	0.90	1.19	1.90
MR	CAP	5	4.01	2.13	4.53	53.04	1.58	2.31	3.32	6.05	7.16
	BN	14	69.27	4.89	23.95	7.07	58.24	66.32	70.23	73.13	75.72
	BNC	1	61.90	-	-	-	61.90	-	61.90	-	61.90
	LF	16	69.25	5.86	34.29	8.46	58.37	64.75	69.06	73.02	81.73
	BC	14	74.93	2.94	8.61	3.92	69.90	72.53	75.58	76.95	80.27
	BCBA	13	64.15	8.23	67.72	12.83	47.82	59.35	64.58	67.02	79.20
	BA	9	33.15	10.46	109.45	31.56	15.00	25.01	32.98	42.18	48.10

Table 2. Normality test for the evaluation attributes.

Variable\Test	Shapiro-Wilk
MR	<0.0001
Al_2O_3 Av.	0.000
SiO_2 Reat.	0.015
Al_2O_3	<0.0001
Fe_2O_3	<0.0001
SiO_2	0.017
TiO_2	0.003

Table 3. Comparison test of the attribute mean values for each lithology presented.

		Lithology					
		CAP	BN	LF	BC	BCBA	BA
Al_2O_3 Av.	%	33.69	29.51	29.51	52.64	53.45	46.84
	GROUP	A/B	B	B	A	A	A
SiO_2 Reat.	%	12.78	10.27	3.33	2.28	3.9	8.91
	GROUP	A	A	B	B	B	A
Fe_2O_3	%	11.35	25.52	48.22	10.18	5.56	3.75
	GROUP	B/C	B/C	A	B/C	C	C
Al_2O_3	%	48.73	40.47	28.47	55.12	58.05	57.18
	GROUP	A/B	B	C	A	A	A
SiO_2	%	14.07	11.24	3.98	2.72	4.65	10.16
	GROUP	A	A	B	B	B	A
TiO_2	%	1.56	1.23	0.89	1.74	1.34	1.04
	GROUP	A/B	B	B	A	A/B	B
MR	%	5.24	67.73	66.67	76.26	64.66	32.17
	GROUP	D	A/B	A/B	A	B	C

SiO_2 and SiO_2 Reat. displayed a normal distribution, while the other attributes did not meet the basic assumptions of normality. The reasons for this are varied, such as the amount of data available that could impact the normality test, the gradational transition between layers, and even geologists' subjective labeling of lithology. However, all attributes were evaluated from an exploratory perspective to assess the potential of image spectroscopy methodology in defining predictive and classification models. In general, there is no predominance in the separation of distinct lithological layers or groups resulting from the mineral content present. For example, the SiO_2 Reat. contents were statistically similar for the CAP, BN, and BA layers and different for BCBA, LF, and BC, while for the Al_2O_3 Av. contents, the mean contents were statistically similar for the BC and BCBA lithologies and different for the groups of lithologies BN and CAP. The same occurred with the other attributes, showing that there is no trend in the average distribution between the lithologies.

As previously described, more than half the samples had low variance in chemistry assay and processing tests. The lack of trends in lithology leads to a conclusion that the prediction models and material separability will be affected by the low variation of mineral concentrations in each lithological layer analyzed.

3.1 Spectral Curves

Figure 9 shows the average spectral curves of the samples sent by Hydro, which refer to the layers of bauxite deposits in the Paragominas exploration region. The distinction between each layer evaluated by the reflectance spectrum is easily observed, both in shape and intensity. In the visible portion (up to 630 nm), differences in the spectral curves are noticeable, highlighting that LF overlays with BC, BCBA is distinct, CAP overlays with BN, and separately, the BA layer has a higher reflectance factor than the others. This range is mainly influenced by the presence of iron oxides (hematite and goethite).

Around 900 nm, the inflection of the spectral curve stands out due to the presence of goethite. At 1 400 nm and 1 900 nm, inflections can be observed resulting from the hydroxyl and water molecules present in 2:1-type minerals such as smectites. In the spectral range of 2 200 nm, the double step characteristic of kaolinite can be observed, as well as the inflection in the range of

2 265–2 275 nm, which is recognized as resulting from the presence of gibbsite. The intertwining that occurs between the spectral curves leads to a conclusion that the prediction and discrimination models may not be as efficient as possible, since part of the models will be affected by similar weightings of bands that make up the model.

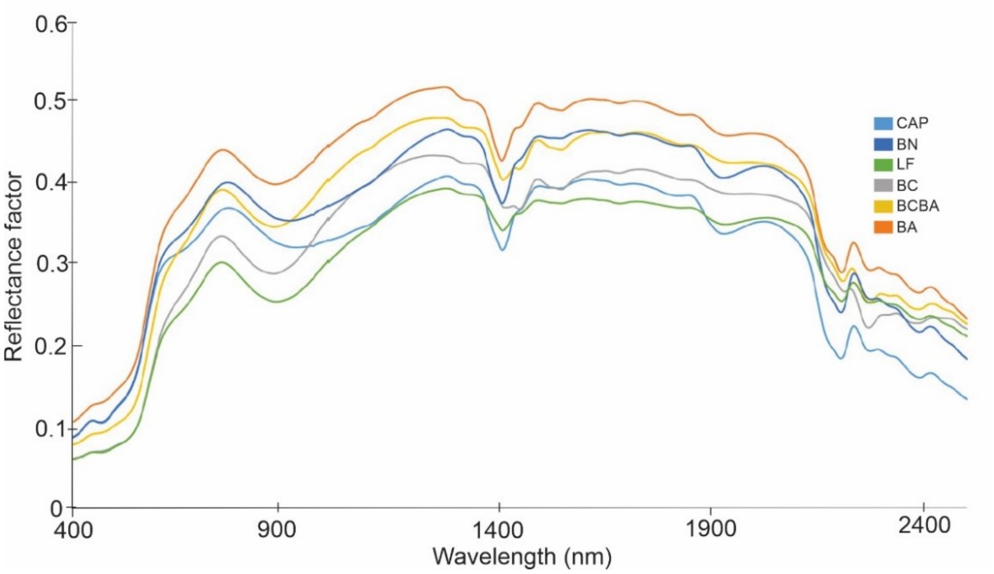


Figure 9. Spectral curves with the SWIR sensor in the FieldSpec system (Laboratory).

The spectral curves of the SWIR sensor were obtained by reading the regions of interest using the ENVI software system (Figure 10). The spectral range of the sensor starts at 1 000 nm. It is observed that the shapes and intensities are similar to those obtained by the Fieldspec sensor, mainly concerning the inflections at 1 400 and 1 900 nm and those referring to kaolinite (2 200 nm) and gibbsite (2 270 nm). The figure also shows the standard deviation (colored band) for each set of samples analyzed. As previously highlighted, there is an overlap between curves belonging to different layers or lithologies. This reflects the heterogeneity of the material, with percentages of constituents that, on average, establish statistical similarities between the layers, as highlighted before.

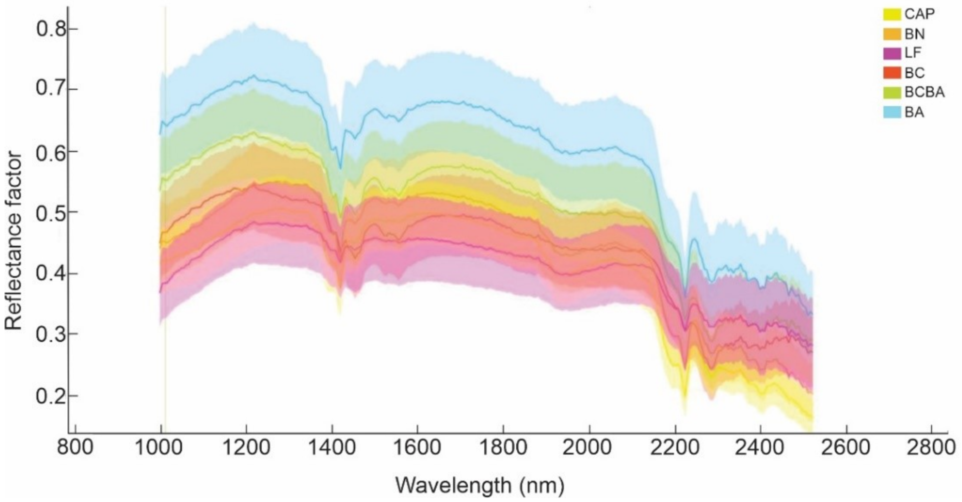
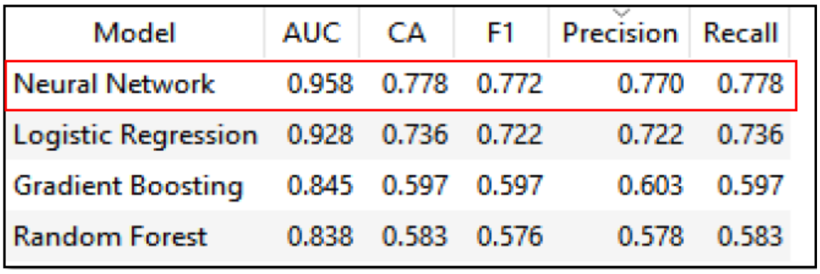


Figure 10. The spectral curves obtained by reading the ENVI system, based on the spectral curves of the SWIR sensor (HyspeX field survey).

The spectral range of the sensor starts at 1 000 nm. It is observed that the shapes and intensities are similar to those obtained by the Fieldspec sensor, mainly concerning the inflections at 1 400 and 1 900 nm and those referring to kaolinite (2 200 nm) and gibbsite (2 270 nm). The figure also shows the standard deviation (colored band) for each set of samples analyzed. As previously highlighted, there is an overlap between curves belonging to different layers or lithologies. This reflects the heterogeneity of the material, with percentages of constituents that, on average, establish statistical similarities between the layers, as highlighted before.

3.2 Discrimination Between Layers (Lithologies) Using Spectral Curves

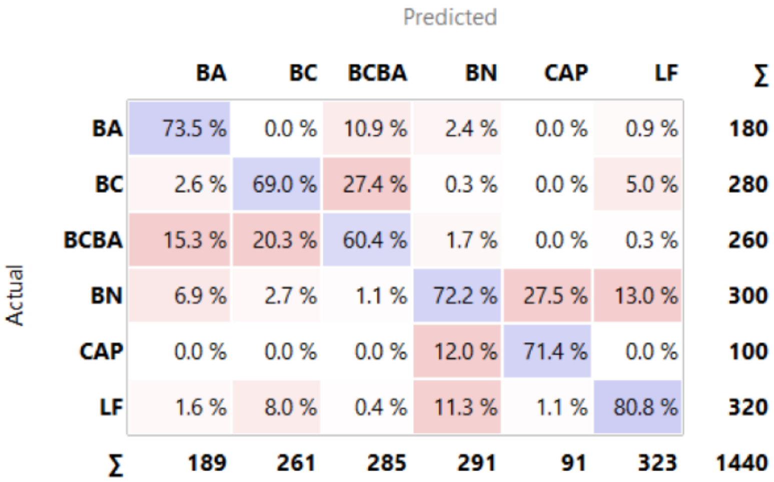
Among the statistical models used for discriminant analysis, the utilization of neural networks yielded the best outcomes (Figure 11). The results of the discriminant analysis can be found in Figures 11, 12, and 13.



Model	AUC	CA	F1	Precision	Recall
Neural Network	0.958	0.778	0.772	0.770	0.778
Logistic Regression	0.928	0.736	0.722	0.722	0.736
Gradient Boosting	0.845	0.597	0.597	0.603	0.597
Random Forest	0.838	0.583	0.576	0.578	0.583

Figure 11. Statistical models performance.

Figure 12 illustrates that the accuracy of samples representing different layers exceeded 60 % (BCBA) and could even reach 80 % (LF). These values were deemed good and very good, indicating the potential for layer separation based solely on the spectral response obtained from the images using the SWIR sensor in the laboratory.



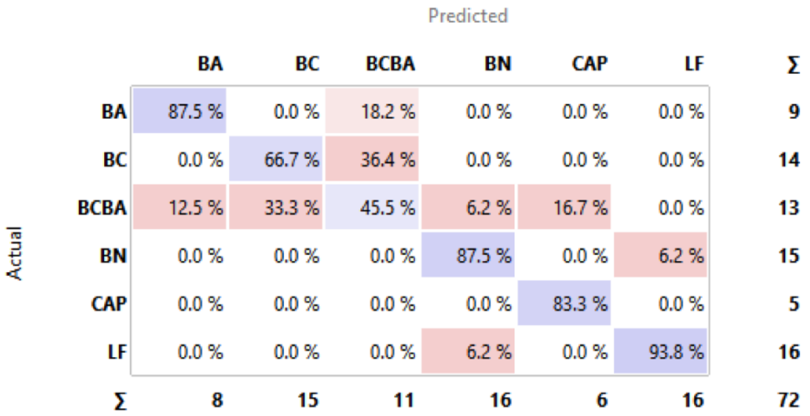
	BA	BC	BCBA	BN	CAP	LF	∑
BA	73.5 %	0.0 %	10.9 %	2.4 %	0.0 %	0.9 %	180
BC	2.6 %	69.0 %	27.4 %	0.3 %	0.0 %	5.0 %	280
BCBA	15.3 %	20.3 %	60.4 %	1.7 %	0.0 %	0.3 %	260
BN	6.9 %	2.7 %	1.1 %	72.2 %	27.5 %	13.0 %	300
CAP	0.0 %	0.0 %	0.0 %	12.0 %	71.4 %	0.0 %	100
LF	1.6 %	8.0 %	0.4 %	11.3 %	1.1 %	80.8 %	320
∑	189	261	285	291	91	323	1440

Figure 12. Confusion matrix for the different layers using SWIR spectroradiometry of bauxite samples.

As previously mentioned, the analysis of variance highlighted similarities between the layers due to the average values of the constituents present. Consequently, we believe that the technique shows promise in terms of distinguishing samples from various collection points (layers or lithologies), helping to reduce similarity errors during lithology logging by geologists.

Similarly, the neural network system proved to be the most effective algorithm for layer separability. The accuracy values for the same layers were higher compared to using only the

SWIR sensor. The BCBA (45.5 %) and BC (66.7 %) layers had the lowest accuracy values, while other lithological sections displayed accuracies above 83 %, reaching 94 % for the LF layer. This trend in accuracy is also evident with the SWIR hyperspectral imaging system (Figure 13). This outcome is primarily attributed to the acquisition of the visible and near-infrared spectral ranges (350–1 000 nm) when utilizing Fieldspec. Since this spectral range predominantly interacts with iron oxides (hematite and goethite), it enables the differentiation of portions with varying levels of these minerals. The LF layer, which has the highest iron concentrations, notably stood out among the evaluated layers, achieving an accuracy of 94 %.



	BA	BC	BCBA	BN	CAP	LF	Σ
BA	87.5 %	0.0 %	18.2 %	0.0 %	0.0 %	0.0 %	9
BC	0.0 %	66.7 %	36.4 %	0.0 %	0.0 %	0.0 %	14
BCBA	12.5 %	33.3 %	45.5 %	6.2 %	16.7 %	0.0 %	13
BN	0.0 %	0.0 %	0.0 %	87.5 %	0.0 %	6.2 %	15
CAP	0.0 %	0.0 %	0.0 %	0.0 %	83.3 %	0.0 %	5
LF	0.0 %	0.0 %	0.0 %	6.2 %	0.0 %	93.8 %	16
Σ	8	15	11	16	6	16	72

Figure 13. Confusion matrix for the different layers using vis/NIR-SWIR spectroradiometry of bauxite samples.

3.3 Estimation of Attributes Using PLS Linear Regression

Table 4 shows that PLS regression successfully estimated MR, Al_2O_3 Av., SiO_2 Reat., Al_2O_3, Fe_2O_3, and SiO_2 with R^2 values above 0.52, reaching up to 0.90.

Table 4. R^2 values for the attributes of bauxite samples using PLS regression.

Method	Al_2O_3 Av.	SiO_2 Reat.	Fe_2O_3	Al_2O_3	SiO_2	TiO_2	MR
HyspeX	0.89	0.83	0.9	0.86	0.8	0.52	0.67
FieldSpec	0.8	0.65	0.77	0.79	0.6	0.55	0.69

The HyspeX imaging sensor range was found to be more effective than the Fieldspec. This is because Fieldspec used a band portion between 350 and 1 000 nm for modeling, which is crucial for identifying iron oxides but overlaps with spectral bands where the inflections of minerals like kaolinite (2 200 nm) and gibbsite (2 270 nm) occur. Therefore, using data within the spectral range of 1 000–2 500 nm led to more robust models for the analysis. Overall, the models performed reasonably well, attributed to the small sample size (72 samples) used to construct the spectral library and the limited variability of its components. As mentioned earlier, there was confusion between the evaluated layers and overlap in spectral responses. It is believed that with a larger sample set, the models will improve.

3.4 Mine Front Classification

Each image was processed in the ENVI software as described in the methodology. Among the classification algorithms used, the support vector machine (SVM) yielded the best results. Regions of interest were selected for each slope and led to supervised classification (Figures 14 and 15).

It was observed that in all evaluated facies, there was a coherent and expected separation of layers. Mine 6 showed classifications in homogeneous bands, which were not as clearly defined for mine 4. Figure 14 also demonstrates collapse in some parts of adjacent slope material, creating difficulty in classification for the selected regions of interest.

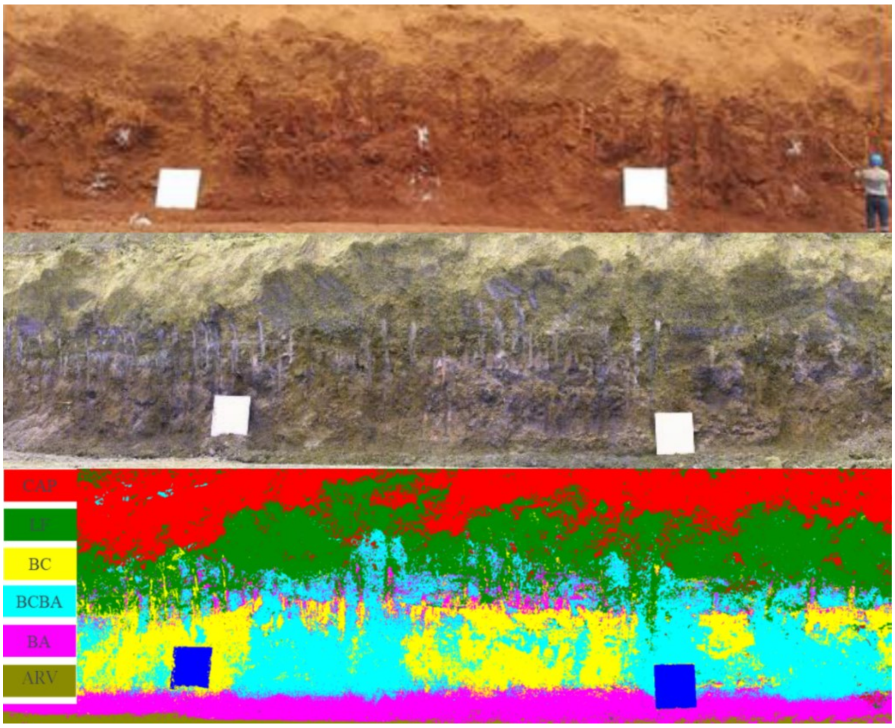


Figure 14. From top to bottom: RGB and false color image obtained by the hyperspectral imager and image classified by the SVM of the mining front located in mine 4.

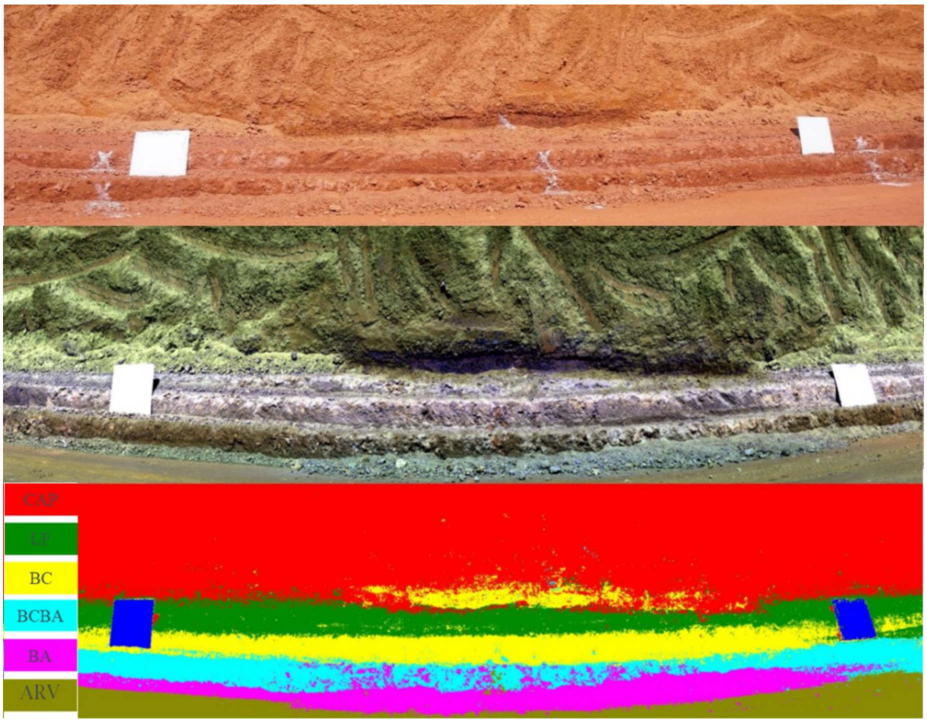


Figure 15. From top to bottom: RGB and false color image obtained by the hyperspectral imager and image classified by the SVM of the mining front located in mine 6.

Analysis of the average spectral curves of the areas showed good separability between the layers, indicating the potential of using spectroradiometry to distinguish layers in bauxite deposits (Figures 16 and 17). The highest reflectance was seen for the CAP layer in both figures. In contrast, the BC and BCBA layers showed the lowest reflectance factor values.

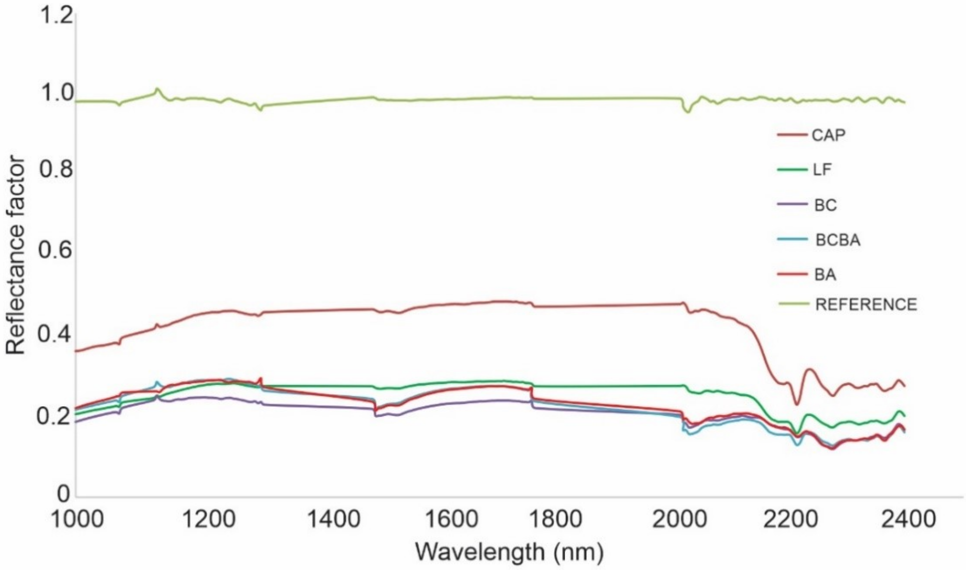


Figure 16. Average spectral curves from mine 4.

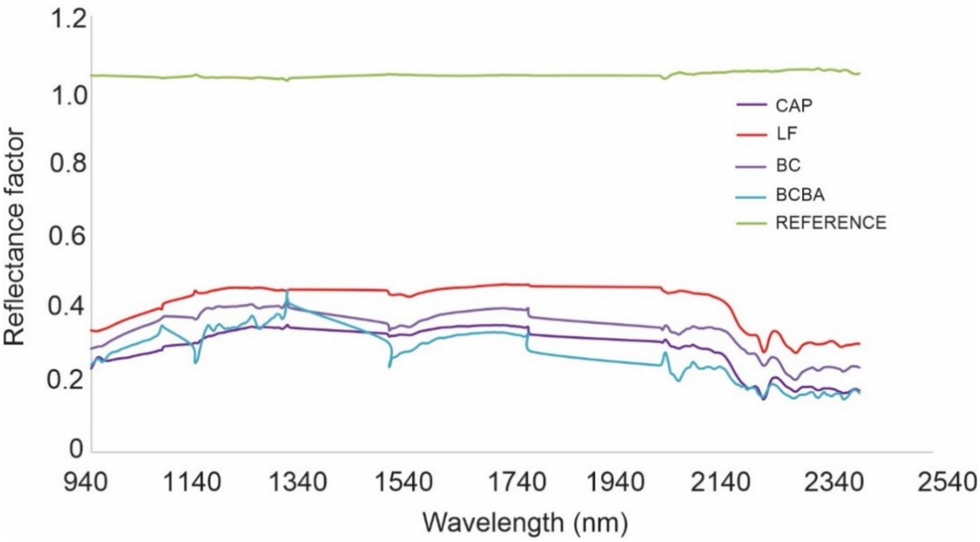


Figure 17. Average spectral curves from mine 6.

4. Conclusions

The results indicated a noticeable difference between the layers assessed through the reflectance spectrum in the laboratory, both in terms of shape and intensity. The utilization of neural networks for prediction yielded the most favorable outcomes (60 to 90 % accuracy), with the spectral range of 1 000–2 500 nm establishing more reliable models for the data analyzed. This underscores the potential for applying hyperspectral methodology in the characterization and quality control of bauxite mining. However, a larger sample size and a more comprehensive spectral library would enhance the refinement and performance of the prediction model, thereby minimizing areas of spectral response overlap.

5. References

1. S.K. Kumari et al., B. Hyperion image analysis for iron ore mapping in Gua iron ore region, Jharkhand, *India. Int. Res. J. Earth Sci.* v. 2 (9), 1-6, 2014.
2. P. Bierwirth, D. Huston and R. Blewett, Hyperspectral mapping of mineral assemblages associated with gold mineralization in the central Pilbara, Western Australia. *Economic Geology*, 97 (4), 819–826, 2002.
3. A. Plaza et al., Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113, 110–122, 2009.
4. G. Vane and A. Goetz, Terrestrial imaging spectroscopy. *Remote Sensing of Environment*, 24 (1), 1–29, 1988.
5. T. Kurz et al., Geological outcrop modelling and interpretation using ground based hyperspectral and laser scanning data fusion. *The international archives of the photogrammetry. Remote Sensing and Spatial Information Sciences*, 37, 1229–1234, 2008.
6. T. Kurz et al., Integration of panoramic hyperspectral imaging with terrestrial lidar data. *The Photogrammetric Record*, 26 (134), 212–228, 2011.
7. S. Monteiro et al., Applying boosting for hyperspectral classification of ore bearing rocks, Grenoble, France, 1, 1–6, 2009.
8. S. Fraser et al., Mineraological face-mapping using hyperspectral scanning for mine mapping and control. In: *6th International Mining Geology Conference*, Darwin, NT, Australia, 21–23 August, pp. 227–232, 2006.
9. R. Murphy, S. Monteiro and S. Schneider, Evaluating classification techniques for mapping vertical geology using field-based hyperspectral sensors. *IEEE Transaction on Geoscience and Remote Sensing*, 50 (8), 3066–3080, 2012.
10. R. Murphy and S. Monteiro, Mapping the distribution of ferric iron minerals on a vertical mine face using derivative analysis of hyperspectral imagery (430–970 nm). *ISPRS Journal of Photogrammetry and Remote Sensing*, 75, 29-39, 2013.
11. Basile Kotschoubey et al., Caracterização e Gênese dos depósitos de bauxita da Província Bauxitífera de Paragominas, Noroeste da Bacia de Grajaú, Nordeste do Pará/Oeste do Maranhão. In: Marini Onildo João et al. (orgs.). *Caracterização de depósitos minerais em distritos mineiros da Amazônia*. Brasília, DF, DNPM-CT Mineral, ADIMB, 2005, 613-698.
12. M. Nanni et al., Partial least squares regression (PLSR) associated with spectral response to predict soil attributes in transitional lithologies. *Arch. Agron. Soil Sci.* 2017, 64, 682–695
13. M. Nanni and J. Demattê, Comportamento da linha do solo obtida por espectrorradiometria laboratorial para diferentes classes de solo. *Rev. Bras. Ciênc. Solo* 2006, 30, 1031–1038.
14. F. Terra, J. Demattê and R. Rossel, Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis–NIR and mid-IR reflectance data. *Geoderma*, 2015.
15. SAS INSTITUTE (Cary, Estados Unidos). SAS software: user's guide, version 6.0. Cary, 1992. 291p.
16. W. Klecka, Discriminant Analysis, Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-019, 1980, Beverly Hills: Sage Publications.
17. Y. Karimi et al., Discriminant analysis of hyperspectral data for assessing water and nitrogen stresses in corn. *Transactions of the ASAE.* 48, 2005,805-813.
18. J. Moura-Bueno et al., Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions. *Geoderma* 337, 2019, 565-581.
19. T. Naes et al., A user friendly guide to multivatiate calibration and classification. Chichester, UK: NIR Publications, 2002

20. R. Dalmolin, Matéria orgânica e características físicas, químicas, mineralógicas e espectrais de Latossolos de diferentes ambientes. 2002. 151f. Tese (Doutorado em Ciência do Solo) – Programa de Pós-graduação em Ciência do Solo, Universidade Federal do Rio Grande do Sul.