# Digital Soil Mapping for the Parnaíba River Delta, Brazilian Semiarid Region

João Victor Alves Amorim ( ✉ amorim@ufpi.edu.br )

Federal University of Piauí - Campus Professora Cinobelina Elvas

Gustavo Souza Valladares

Federal University of Piauí

Marcos Gervasio Pereira

Federal Rural University of Rio de Janeiro – Campus Seropédica

Mirya Grazielle Torres Portela

Federal University of Piauí - Campus Amílcar Ferreira Sobral

Andréa Maciel Lima

Universidade Federal de Goiás

# Abstract

Soil mapping is a permanent demand, but the traditional method does not allow fast execution and low cost. Digital soil mapping (DSM) aims to improve the process by working with models that treat soil spatial variability quantitatively. In this perspective, the objective of the study is to perform DSM of the Parnaíba River Delta, Northeastern Brazil, through the decision tree (DT) integration technique using a set of attributes derived from the digital elevation model (DEM) and satellite images as input parameters. Data matrices were created considering different soil groups. The performance of the J48 machine learning algorithm (DT) was assessed for a set of 2 data matrices. From the data processing, digital soil maps were created and validated by means of error matrices, whose reference points were classified in the field, and validated by means of a pre-existing traditional soil map of the area. The results revealed that the attributes derived from satellite image stood out from those derived from DEM in the prediction of soil groups. Based on the validation coefficients applied (overall accuracy, Kappa index, user's accuracy and producer's accuracy), the quality of the classification was satisfactory, despite the complexity of the environment.

# 1. Introduction

Digital Soil Mapping (DSM) is a tool capable of increasing the feasibility of performing soil surveys, by using information related to geology, relief, hydrography, climate, vegetation, among others, for classification. It aims to optimize time and resources, as well as improve conventional work of pedology, enabling larger areas to be mapped (Nolasco-Carvalho et al., 2009).

The application of pedometry techniques depends on the objective and on the detail intended for mapping, whereas the final use of soil survey information determines the required accuracy. Consequently, these cannot be applied to any situation without considering the specific needs and convenience of the hypotheses inherent to each technique (McBratney et al., 2000).

There is still no systematized knowledge about soil spatial variability and, in particular, for DSM there is still no standard of work that is appropriate to the various environments. However, given the growing popularity of DSM, the number of models and procedures for its optimization of performance in modeling increased (Minasny and McBratney, 2015). Some models applied to soil mapping include: geostatistics (Kempen et al., 2012); fuzzy logic (Nolasco-Carvalho et al., 2009; Taghizadeh-Mehrjardi et al., 2015; Rizzo et al., 2016); C.L.O.R.P.T. techniques based on environmental correlation, application of artificial neural networks (ANN), classification and regression trees (CART) and regression models (Silveira et al., 2013; Brungard et al., 2015; Heung et al., 2016; Chagas et al., 2010), in addition to models that apply decision trees (DT) (Höfig et al., 2014).

The approach by DT has been used for having the advantages of enabling the explicit expression of soil-landscape relationships, grouping and search for patterns, besides enabling the understanding of how these data are interrelated (Kheir et al., 2010).

In this context, this study aims to perform the digital mapping of soils in the Parnaíba River Delta area, coastal zone of the state of Piauí - Brazil, using the technique of data mining by DT. The mapping will be the final product of the processing of predictive variables from the digital analysis of the relief, as well as attributes derived from orbital sensor images.

## 2. Materials And Methods

## 2.1 Study area:

The study area is located in the northern portion of Piauí's coastline and in the septentrional northeast of Brazil, comprising part of the Parnaíba River Delta Environmental Protection Area (created by the Federal Decree of August 28, 1996) and part of the Parnaíba River Delta Marine Extractivist Reserve (created by the Federal Decree of November 16, 2000), more precisely in the region delimited by the Igaraçu River to the southeast, Parnaíba river to the west and the Atlantic Ocean to the north, covering the municipality of Ilha Grande, and part of the municipality of Parnaíba. The study area is approximately 282 km$^2$.

The area in question deserves attention for its socioeconomic and environmental importance for the State, given the diversity of environments and natural systems that can be found there (Portela et al., 2020) and that support important economic activities, such as tourism, extractivism, among others. It is regulated in Brazilian legislation by the Federal Law N° 9,985 of July 18, 2000, which establishes the Brazilian National System of Nature Conservation Units (SNUC—acronym in Portuguese) (Brazil, 2000). It classifies conservation units into full protection and sustainable use units, and the two aforementioned units are classified as sustainable use units (Fig. 1).

With regard to geo-environmental aspects, the area has Tropical-Equatorial climate, with seasonally distributed rainfall, recording maximum levels in autumn, and it is possible to identify two well-defined seasons: a rainy season during the first months of the year, from January to May, and a dry that occurs in the second half, from July to December (Mendonça and Danni-Oliveira, 2007).

In 2016, the study region had water surplus, which extended until July, and months without accumulated precipitation, with monthly average temperature reaching 35°C in September. Compared to the climatological normal of precipitation in 30 years (Fig. 2), the precipitation in this period was lower than that commonly observed, because the northeast of Brazil was affected by a severe drought between the years 2012 and 2015, which caused impacts not seen in previous decades (Marengo et al., 2017).

According to geology, sediments from the Quaternary period predominate in the Parnaíba River Delta. These are represented by the beach, wind, marine and lagoon deposits and alluvial-colluvial deposits. The area is subdivided into seven geological units, namely: delta and fluvial channels; coastal deposit of recent beaches; mobile coastal wind deposits; fixed coastal wind deposits; alluvial-fluvial deposits; deposits of swamps and mangroves; and sandy deposits (Valladares and Cabral, 2017).

The geomorphological units that make up the study area comprise the aggradation reliefs. This type of relief is one with predominance of depositional processes, both of the continental type, as is the case of the fluvial plain, and of the coastal type, as is the case of the coastal plain, the river-marine plain, the marine alluvium-colluvium plains and the fluvial-lagoon plains, as well as mobile dunes, sandspits and beaches. Through the mapping carried out by Sousa (2015), at a scale of 1:100,000, it is possible to identify nine geomorphological units in the Parnaíba River Delta area, namely: sandspit; delta and fluvial channels; stabilized dunes; mobile dunes; paleodunes; beach; wind plain; marine fluvial plain; and fluvial plain and terrace.

For being a conservation unit of sustainable use, classified by the SNUC, the study area has part occupied by urban spaces, agricultural activities and wind energy production, with restricted access areas. The region has a great plant biodiversity translated into the various types of vegetation, which predominate in the landscape and mix with other formations such as mangroves, coastal *restingas*, and vegetation of Cerrado and Caatinga.

According to Costa and Cavalcanti (2010), the vegetation present in the area is constituted mostly of woody species, characterized by trees spaced with irregular crowns and secondary aspect, alternated with subsistence agriculture. The vegetation has *restinga* physiognomies of flooded and non-flooded scrubland around tree species, these being demarcated by areas of depression, resulting from wind activity on the dunes, flooded in the rainy season, as well as formations of fields and carnauba stands (Santos-Filho et al., 2010). In addition, there is also the presence of secondary physiognomy vegetation with significant interpenetration of typical Caatinga species, unique in the world, besides having mangroves with high plant biomass (Portela et al., 2020).

## 2.2 Obtaining predictive attributes

The attributes used were those referring to geomorphometry (topographic attributes) derived from the Digital Elevation Model (DEM) of the area and those from remote sensing products. The DEM was generated from an SRTM image of the area and extraction of vector planialtimetric data from the topographic sheet of Geographic Service Directorate of the Brazilian Army (Brazil, 1972) Parnaíba sheet (SA 24 Y-A-IV) at a scale of 1:100,000. In geographic information system (GIS), the variables hypsometry, slope, curvature, flow direction, flow accumulation and topographic wetness index were generated.

The slope was generated from the DEM using two local finite differences in the x and y directions (Horn, 1981). The slope classes delimited were: 0 to 3% (Flat), 3 to 8% (Gently Undulating), 8 to 20% (Undulating), 20 to 45% (Strongly Undulating), 45 to 75% (Mountainous), above 75% (Rugged).

Curvature is the result of the combination of horizontal (convergent, planar or divergent) and vertical (concave, rectilinear or convex) curvature classes (Valeriano, 2008). For the discretization of the curvature in the histogram, values greater than 0.5 were used to represent concave-convergent slope segments, values lower than − 0.5 for convex-divergent segments and values from − 0.5 to 0.5 for rectilinear-planar segments.

The water flow direction in the drainage network is a regular grid defining the flow directions, based on the line of greatest slope of the terrain. It allows the observation of the direction of water flow in the slopes, in addition to the visualization of the relief (Rennó et al., 2008).

The cumulative flow was obtained by using the previously determined direction of flow, indicating the degree of confluence of flow, and it may be associated with the ramp length factor applied in two dimensions (Guedes and Silva, 2012). According to Sobrinho et al. (2010), flow accumulation represents the hydrographic network, and it is possible to assemble a new grid containing the values of water accumulation in each pixel. Thus, each pixel receives a value corresponding to the number of pixels that contribute for the water to reach it.

The topographic wetness index, according to Sirtoli (2008), is used to characterize the spatial distribution of surface saturation zones and water content in landscapes. For Moore et al. (1993), this index further demonstrates the effects of relief on the location and extension of water accumulation areas. For Silveira et al. (2013), it has application for separating soils with hydromorphic character, which occur in areas of flat relief, from other classes of soils that occur in areas of relief ranging from flat to gently undulating. It is defined as a function of slope and the contributing area per unit width orthogonal to the flow direction (Chagas, 2006) obtained through Eq. 1:

$$TWI = \ln \left( \frac{As}{tan\beta} \right) \qquad \text{Equation 1}$$

where: $As$ is the contributing area multiplied by the size of the grid cell by m$^2$; and $\beta$ is the slope of the cell.

Regarding the attributes derived from orbital images, it was decided to use Landsat 8 OLI/TIRS C1 Level-2 (U.S.G.S. Earth Explorer) images, acquired on October 14, 2017. Bands 2 (blue), 4 (red) and 6 (shortwave infrared) of the Operational Land Imager (OLI) sensor were used separately. The products have spatial resolution of 30 meters, subjected to orthorectification and radiometric corrections and are under the coordinate system UTM Dtuna WGS 1984, zone 24 South.

According to Barsi et al. (2014), band 2 is used more effectively in bathymetric mappings, distinguishing soil from vegetation and deciduous vegetation from tree vegetation. Band 4 discriminates vegetation through the chlorophyll absorption of healthy green vegetation and is also useful to delimit the limits of soil classes and rock types. Band 6 discriminates the moisture contents of soil and vegetation, being sensitive to turgor or amount of water in plants (Jensen, 2009).

In addition to those mentioned above, the Thermal Infrared Sensor (TIRS) of Landsat 8, orbit/point 219/062, known as thermal band, in the infrared range, with wavelength from 10.6 to 11.2 m, was also used, without performing any transformation in the image. The spatial resolution of the bands of the TIRS sensor onboard the satellite is 100 meters, but they are resampled to 30 meters in the data product delivered.

Indices obtained from the relationship between different bands of this sensor were also used. The indices used were clay minerals (CLAY), obtained by the division of band 6 (near infrared) (1.57–1.65 μm) by band 7 (near infrared) (2.11–2.29 μm), and iron oxide (IRON), obtained by the division of band 4 (red) (0.64–0.67 μm) by band 2 (blue) (0.45–0.51 μm) (Chagas et al., 2010), originally developed for Landsat's TM sensor and according to equations 2 and 3:

$$CLAY = \frac{BAND6}{BAND7} \quad \textbf{Eq. 2}$$

$$IRON = \frac{BAND4}{BAND2} \quad \textbf{Eq. 3}$$

The Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Water Index (NDWI) were also used. For Chagas (2006), vegetation indices are formed from the combination of several spectral values that are summed, divided or multiplied to produce a single value that indicates the amount or vigor of vegetation within a pixel. Thus, for the Landsat 8's OLI sensor image, the NDVI index was obtained by Eq. 4:

$$NDVI = \frac{BAND6 - BAND4}{BAND6 + BAND4} \quad \textbf{Eq. 4}$$

where, BAND 6 = near infrared and BAND 4 = red.

The NDWI was designed with the purpose of delimitating open water environments, automating the determination of the threshold between water and land (terrestrial vegetation and soils). In this perspective, the equation proposed by equation Gao (1996) was applied to highlight the water bodies, using reflectance data of the bands corresponding to 860 nm and 1240 nm, respectively (Eq. 5). The reason for this choice lies in the spectral pattern of water. In the range close to 860 nm, the expected reflectance of water is very low and in the range of 1240 nm, the expected reflectance is zero:

$$NDWI = \frac{BAND5 - BAND6}{BAND5 + BAND6} \quad \textbf{Eq. 5}$$

To apply Eq. 5 in images acquired by Landsat 8, the reflectances of bands 5 (850 nm-880 nm) and 6 (1570 nm-1650 nm) were chosen.

# 2.3 Creation and analysis of the data matrix for DT composition

For the mapping, 119 points were sampled in the field. Of these, 98 were used for model training and 21 for subsequent map validation. The location of the sampled points is shown in Fig. 3. The samples were stratified and drawn within each mapping unit established in the conventional soil map (Cabral, 2018) (Fig. 3). The division was carried out in order to keep both sets well distributed throughout the study area and in proportions representative of all units.

Thus, two data matrices were determined, which consisted of the union of all the generated attributes, always using the same combinations of predictor variables. The matrices were created for the mapping units of the pre-existing conventional soil map and for a set of associations determined based on the landscape features of the study area with close correlation with the existing soils, mainly due to the parent material. Table 1 shows the details of the data matrices created for predictive mapping.

It is important to highlight that, to construction of the matrices, it was necessary to make the class AReu2 of the conventional mapping of Cabral (2018) compatible, by dividing it into MD (mobile dunes) and SB (sand beach), so that there was no erroneous interpretation by the software when classifying these areas as sandy texture soils.

Table 1
– Description of data matrices created for DSM

| Matrices | Predictive attributes | Classes |
|---|---|---|
| 1) Map units adapted from Cabral (2018) | All parameters: hypsometry, slope, curvature, flow direction, flow accumulation, TWI, band 2, band 4, band 6, thermal band, IRON, CLAY, NDVI, NDWI. | GLeu (GLeu + SCgl + GLum) |
| | | GLti |
| | | AReu (AReu + PZcb) |
| | | SNap (SNap + FLeu + CMfv + VRso) |
| | | MD – Mobile dunes |
| | | SB – Sand beach |
| 2) Landscape Features | All parameters | GL – Soils with gleyc characteristics |
| | | AR – Soils with sand texture |
| | | FL – Soils with fluvic materials |
| | | MD – Mobile dunes |
| | | SB – Sand beach |

Legend: TWI: topographic wetness index, IRON: iron oxide index, CLAY: clay minerals index.

The databases were exported as a table and converted to a comma-delimited file (CSV – MS-DOS) and then rearranged to the Attribute-Relation File Format (ARFF), which describes a list of instances that share a set of attributes, for subsequent use in Weka 3.8 software (Witten and Franck, 2005).

In Weka, for analysis of the data sets, the performance of the J48 classifier algorithm, corresponding to the DT method, was verified. The result of the processing of each data matrix was used for the mapping through the prediction of soil classes by the algorithm, which automatically generates an accuracy assessment.

In all procedures performed, the method known as K-fold Cross-Validation was used as a technique for stratifying the database into training and test sets (Dias et al., 2016). For Witten et al., (2011), this technique has become a standard for applications using data mining methods. Furthermore, according to

studies evaluated by these authors, number 10 (ten) should be adopted as the standard value for the number of partitions of the data (K).

## 2.4 Generation of digital soil maps and accuracy assessment

From the procedures described, one digital soil map was generated for each data matrix. The result of the data matrix with best performance in classification was exported to text file (.txt) and processed in the conversion software ADtoSIG (Ruiz et al., 2011). This program converts the file to a possible base to be interpreted in GIS. The ADtoSIG output file was used in the ArcGIS Raster Calculator tool. The map was produced from the conditional rules established by the program and the raster format files of the predictive variables. In this case, the choice of which attribute to use was determined by the classifier algorithm during classification.

The accuracy of the maps generated by pedometry for the study area was evaluated by the validation through the 21 points classified in the field for validation and through the comparison between the maps generated with the pre-existing conventional soil map (Fig. 3). These data were compared with the maps predicted by means of error matrices and, from the results, the following measures of accuracy were calculated: kappa index, overall accuracy, user's accuracy and producer's accuracy. Qualitative interpretations of the kappa coefficient were based on ranges that represent image quality, according to Landis and Koch (1977) (Table 2).

**Table 2 –** Image quality classification according to kappa index ranges.

Source: Adapted from Landis and Koch (1977).

| Kappa coefficient | Image quality |
|---|---|
| Less than 0 | Very poor |
| 0.01−0.20 | Poor |
| 0.21−0.40 | Reasonable |
| 0.41−0.60 | Good |
| 0.61−0.80 | Very good |
| 0.81−1.00 | Excellent |

## 3. Results And Discussion

From the DEM, six geomorphometric attributes were generated for the study area: hypsometry, slope, curvature, flow direction, flow accumulation and topographic wetness index. For hypsometry, the values

varied from 0 to 5 meters above sea level, and the ranges from 10 to 15 and >15 meters were limited to the areas of mobile dunes. Regarding slope, 63.5% of the area has flat relief (0 to 3%). Analysis of the curvature shows that there is a predominance of rectilinear-planar segments, which correspond to 94.74% of the area. The convex-divergent and concave-convergent areas together correspond to only 5.26%, occurring in the ranges of mobile dunes.

Flow direction shows a predominance of the North (22.26%) and West (19.47%) classes, justified by the fact that the area is located in the low course of the Parnaíba River, mouth with the Atlantic Ocean, general base level. Regarding flow accumulation, sites with high-value cells indicate areas with high viability for the occurrence of drainage. In the Parnaíba River Delta, the highest values are found in areas of occurrence of hydromorphic soils and with fluvic character, to the southeast, formed under strong influence of alluvial sediments.

For the topographic wetness index, high values were observed for more saturated areas (lowland areas that accompany the river plains or areas that favor water accumulation in the soil) and lower values were observed for well-drained areas and areas with slope greater than 8% (mobile dune fields), with a variation from − 1.5 to 13.9. It is worth pointing out that the equation for obtaining TWI is Ln ($a$ / tan ($B$)), where $a$ is the 'specific' catchment area (that is, the upstream inflow area normalized for a measure of contour length/flow accumulation) and $B$ is the slope gradient, in radians, in the grid cell (slope). TWI will display negative values when the catchment area is lower than tan(B), since the natural logarithm of any value lower than 1 will be negative. For the Parnaíba River Delta, flow accumulation and slope had zero values for most of the area, justifying the variation of the values obtained for TWI. Figure 4 shows the frequency histogram of these attributes.

From the images of Landsat 8's OLI sensor, it was possible to generate the following predictive variables: band 2 − blue, band 4 − red, band 6 − medium infrared, band 10 − thermal band, CLAY, IRON, NDVI and NDWI. Band 2 showed high values of reflectance in the range of mobile dunes and in the areas of sandspit and beach, assisting in the identification of sandy soils and soils with hydromorphic character, represented by the low values of digital numbers.

Band 4 proved useful for discriminating the vegetation, as it is the red band of absorption by chlorophyll. Like band 2, band 4 showed high values of reflectance for areas of sandy soils. In sites where the vegetation is dense, such as mangrove forests and areas of swamp vegetation, there was greater absorbance in the spectral range of red. Band 6 (shortwave infrared), analyzed individually, also showed satisfactory results for a differentiation of the dune fields with the sandspit and beach areas, where the values were high. Through the interpretation of the low values it was possible to delimit the mangrove areas. Although the soils found in these sites are similar in character, these differentiations were useful in the training of the prediction model used.

The thermal band of the TIRS sensor makes it possible to remotely estimate soil temperature by transforming the digital number of the image into radiance. For the study object, pixel values ranged from 22.9 to 37.5 and were effective in the delimitation of the areas of Fluvial-Marine Plain (values between

24.1 and 25.4), being possible to find associations of hydromorphic soils (Thionic Gleysol + Umbric Gleysol + Gleyic Solonchack + Thionic Histosol), all of clayey or indiscriminate texture. The image also assisted in the identification of areas of fluvial-alluvial deposits (values between 31.2 and 37.5), representing associations mainly of soils originated from alluvial sediments with indiscriminate texture (Abruptic Solonetz + Fluvic Cambisol + Eutric Fluvisol + Sodic Vertisol).

CLAY showed digital values ranging from 0.43 to 3.0. Higher values were associated with the presence of clay minerals present in soils of areas of fluvial-marine plain and fluvial terrace, which can be correlated with hydromorphic soils. The lowest values represent soils with the presence of primary minerals, mainly sands constituted by quartz found in the areas of mobile dunes, sandspit and beach.

IRON ranged from − 22.2 to 34.1, and it was possible to identify higher values in areas of mobile dunes and fluvial-alluvial deposits, with associations of soil with sandy texture and alluvial soils with indiscriminate texture, which contain considerable amounts of primary minerals. Lower values indicate deposit areas of swamps and mangroves, suggesting the occurrence of hydromorphic soils. Similar values were verified for the areas of stabilized dune and paleodunes. In these places there is a predominance of herbaceous and/or shrubby plant species and it is common to find water bodies due to the low position in the relief and proximity of the groundwater to the surface, where Arenosols can occur and with strong restriction to drainage, that is, the presence of hydromorphism. This situation of relief causes reduction of iron and manganese oxides to their most soluble ionic forms.

For NDVI, the digital values of pixels between − 1 and − 0.10 indicated water bodies. The pixels with values from − 0.09 to 0.15 indicate areas with low vegetation response. These areas correspond to Mobile Dunes, Sandspit and Beach, consisting predominantly of poorly consolidated sandy materials. Values between 0.15 and 0.30 represent areas with presence of herbaceous and/or shrubby undergrowth and, associated with it, mainly Eutric Aerosol. The values between 0.3 and 0.4 classified the areas of open natural fields with the presence of Eutric Fluvisol. Values from 0.4 to 0.6 indicate areas with predominance of more shrub/tree sized vegetation, where it is possible to find associations of Abrupt Solonetz + Fluvic Cambisol. Values from 0.6 to 1.0 indicate the areas of fluvial-marine plains, with dense mangrove tree vegetation and swamp vegetation, as well as occurrence of hydromorphic soils.

NDWI is highly correlated with the water content in the vegetation cover and makes it possible to monitor changes in biomass and evaluate the water stress of vegetation. For the study area, the variation was from − 0.08 to 0.87, indicating the presence of areas with large water accumulation, and it was possible to find in these sites hydromorphic soils, both Gleysols in areas of fluvial-marine plain (range between 0.06 and 0.18) and Fluvisols in areas of fluvial terrace (range between 0.18 and 0.26). The values between 0.26 and 0.34 indicated areas of fluvial-alluvial deposits with vegetation of the lowland tropical forest type with the presence of Carnauba palms (*Copernicia prunifera*), where Abruptic Solonetz and Fluvic Cambisol occur. Between 0.34 and 0.45, it was possible to classify the wind plain, indicating the occurrence of sandy soils (Eutric Arenosol and Carbic Podzol). It is worth pointing out that the orbital image was obtained in October, coinciding with the dry period (low rainfall). In the first quarter of the year,

water concentration is common in these places. The areas of sandspit, beach and mobile dunes are identified by the values between 0.45 and 0.87. Figure 5 shows the frequency histogram of the spectral attributes.

By crossing the predictive variables with soil classes, mapping units and landscape features, it was possible to form 2 data matrices, previously described. After creating the data matrices, the analyses were performed by the Weka program, and the cross-validation option with 10 partitions was selected to test the J48 algorithm through the accuracy assessment. The result of the hit rates of the classifier algorithms for each data matrix is presented in Table 3. Processing generates an automatic kappa value for each tested set.

Table 3
– Results of data matrices processing for MDS

| Matrices | CRI (%) | Kappa |
|---|---|---|
| 1) Map Units (CABRAL, 2018) | 56.1 | 0.45 |
| 2) Landscape Features | 75.5 | 0.66 |
| Legend: CRI - Correctly Ranked Instances | | |

For matrix 1, the classifier model selected the parameters NDVI, IRON, thermal band, band 2, hypsometry and flow accumulation as the ones with greatest predominance for the classification. Based on the confusion matrix, it is observed that of the seventeen cells previously indicated as GLti, fourteen were correctly classified, with a hit percentage of 82.3%. Only half of the cells indicated as GLeu were correctly classified, and of the total, 37.5% were classified as SNap. A justification is that the SCgl present are located in environments similar to those of FLeu, representative of the SNap class. The SNap class show hits in seventeen cells out of a total of thirty (57%), with 20% of the remaining cells classified as Gleyc soil associations (Fig. 6).
Based on the classifier model of matrix 2 (Fig. 6), it is verified that the parameters used by the algorithm were: NDVI, IRON, thermal band, band 2, CLAY, band 4, hypsometry and curvature. NDVI was decisive for the construction of the model, being able to individually separate two classes, SB and MD, with the help of IRON. Another attribute of great representativeness was the thermal band, due to its ability to determine soil moisture, which identified the GL class.

For the determination of the other classes, the other parameters were considered in the classification. However, it is worth mentioning that the geomorphometric attributes had lower relevance, and only hypsometry and curvature were entered in the model. This information, combined with the result obtained from the analysis of the matrix 1 model, indicates that the attributes obtained from the DEM, for the study area, have low effectiveness for classification, due to the low variability. Another limitation refers to the

spatial resolution of the DEM employed, so a product with higher level of detailing is necessary to verify greater correlation of morphometric attributes with soils.

The digital soil maps of matrices 1 and 2 were generated and validated by means of an error matrix that confronts the classes predicted by the model with the actual classes identified in the field at the 21 points pre-selected for this validation. The maps were also compared with the pre-existing conventional soil map, according to the procedures described above. Figure 7 illustrates the digital soil map of the Parnaíba River Delta generated through the data matrix 1.

For this processing, the largest class was AReu, which comprises an area of 89.28 km² (38.57%). In this class the Eutric Arenosols stand out, with simple grain structure. They contain the mineral quartz as predominant in all of their fractions. This class also comprise Carbic Podzol, sandy soils that have spodic horizon, with illuvial accumulation of humified organic matter, combined with aluminum, with very dark gray color (10YR 3/1) in the subsurface diagnostic horizon and black color (7.5YR 2.5/1).

The second largest class was SNap (42.11 km², 18.19%). Among the soils, the occurrence of Abruptic Solonetz stands out; these are soils with high sodium saturation, with prismatic or columnar structure. Its high textural gradient causes great susceptibility to erosion, also favored by the low permeability of B horizon. There is also the occurrence of Eutric Fluvisols, which are poorly evolved mineral soils formed from recent alluvial deposits. In addition to these, there are also Fluvic Cambisols, which consist of mineral material with cambic horizon, underlying the A horizon, with not very advanced degree of development and irregular variations of granulometry in subsurface. The last component of the association is the Sodic Vertisol, which consists of mineral material, with vertic characteristics in the horizons, appearance of cracks in the dry period due to clay shrinkage and swelling, and slickensides.

The grouped classes GLeu and GLti cover an area of 59.46 km² (corresponding to 25.68%). In this unit, the soils found have gleyic properties. These soils often have mottles or variegated colors and can assume any hues and values as long as the chroma is less than or equal to 2.

Classes MD and SB extended for 27.59 km² (11.91%) and 37.99 km³ (16.41%), respectively. Mobile dunes are large, individual moving masses of sand, consisting of simple and/or composite wind dunes, as well as large strips of sand stretched near the beach line. The sandspits and beach are accumulations of sand located between the base of the modal waves and the boundary of the beach, deposited mainly by the waves, but are also influenced by the tides and the local topography.

Regarding the accuracy of the mapping, the validation by means of an error matrix, expressed in Table 4 showed overall accuracy of 61.90%, higher than that of the previous map. The same is true for kappa, which for this map reached 0.52, a good agreement according to criteria of Landis and Koch (1977).

It is worth mentioning that, according to Ten Caten (2011), the mean value of kappa index in studies conducted in Brazil is 0.47. This value is higher than that reported in the international literature, for example, 0.39 for flat areas, reported by Hengl and Rossiter (2003).

The map generated from data matrix 2 is represented in Fig. 8. Based on the image, it was verified that the most comprehensive class for the study area was sand texture soils - AR, with about 90.31 km² (39%). The second most comprehensive class was that of soils originated from alluvial sediments, of indiscriminate texture - FL, with 60.24 km², corresponding to 26% of the area.

The class of soils with gleyc characteristics – GL represents 19.4% of the area (approximately 44.91 km²). Most of these soils are representative of the fluvial-marine plain areas (mangrove forest areas) that are subjected to strong influence from the Parnaíba River, its tributary, the Igaraçu River, and the Atlantic Ocean. Classes MD and SB extended for 22.99 km² (9.9%) and 13 km³ (5.6%), respectively.

The error matrix of the digital map generated from matrix 2 is found in Table 5. The overall accuracy was 67% and the kappa index was 0.57, indicating good agreement according to the adopted criteria for evaluating the accuracy.

In general, there was good agreement with regard to the accuracy of the digital maps. The J48 algorithm model reached 56.1% and 75.5% accuracy when data matrices 1 and 2 were analyzed, respectively. These models were used to generate the maps that, when confronted with the points of validation through the error matrix, showed proportions of correctly classified samples of 61.90% and 67%, values that are positively correlated and with similar quantities.

The accuracy of the model depends on the way in which the data matrices were created, on the preprocessing procedures employed in them, and on the ability of the algorithm to interpret and extract patterns and associate them with the different soil classes.

Aiming at a more effective comparison between mapping methods, a validation was also performed from the pre-existing traditional map of Cabral (2018) (Fig. 3). For the map generated from data matrix 1, the overall accuracy was 55% and kappa reached 0.40, according to the error matrix shown in Table 6. It was necessary to make the legend of the digital maps compatible for effective comparison with the traditional map, where the classes MD and SB, corresponding to the class AReu2, were grouped. The result of this crossing is illustrated in Fig. 9, where it is possible to verify the areas of agreement and disagreement between the two mappings.

The map generated based on matrix 2, when validated by the traditional map, showed overall accuracy of 72.64% and kappa of 0.61, indicating good agreement in the classification. Table 7 shows the error matrix of the mapping.

The error matrices of digital maps with legends expressed in mapping units (or other association criteria) show that overall accuracy and kappa are higher compared to maps with legends expressed in a single soil group, despite using the same set of predictive variables.

# 4. Conclusions

The classification algorithm used, J48, showed satisfactory results at the mapping scale employed, demonstrating great potential to support soil cartography, since the algorithm itself selects and defines

the priority of each attribute during classification. When compared with the pre-existing map, the maps generated by DT showed significant agreement, indicating that the method can be extended to other coastal environments.

As for the creation of the data matrices, it is verified that the potential for use of the DEM SRTM data proved to be low for the area, mainly due to imperfections that reduce the accuracy of the entire data set. Variables derived from the spectral image were more adequate when the classifier models of the two matrices were considered.

# Declarations

## DATA AVAILABILITY STATEMENT

The datasets generated during and/or analysed during the current study are not publicly available because they are part of a research that it's not completely published yet but are available from the corresponding author on reasonable request.

## DECLARATION OF COMPETING INTERESTS

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.

# References

1. Barsi, J.A., Lee, K., Kvaran, G., Markham, B.L., Pedelty, J.A. (2014). The Spectral Response of the Landsat-8 Operational Land Imager. Remote Sensing 6, 10232–10251. https://doi.org/10.3390/rs61010232.
2. Brazil (2018). NATIONAL INSTITUTE OF METEOROLOGY OF BRAZIL – INMET. Climatological Normal Precipitation (1981/2010). Brasilia - DF.
3. Brazil (2000). Law No. 9,985 of July 18, 2000. Regulamenta o art.225, § 18, incisos I, II, III e IV da Constituição Federal, institui o Sistema Nacional de Unidades de Conservação da Natureza e dá outras providencias.

4. Brazil (1972). Brazilian Army. Geographic Service Directorate. *Topographic sheet Parnaíba, Folha SA 24 Y-A-IV*, scale 1:100,000.

5. Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, A.S., Edwards Jr, T.C (2015). Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239–240, 68–83. https://doi.org/10.1016/j.geoderma.2014.09.019.

6. Cabral, L.J.R.S (2018). *Levantamento pedológico da Planície do Delta do Parnaíba, PI*. Master's degree dissertation, Federal University of Piauí, Teresina, Brazil.

7. Chagas, C. S. (2006). *Digital Mapping of Soils by Environmental Correlation and Neural Networks in a Watershed in the Sea Hills Area*. Ph.D. Thesis, Federal University of Viçosa, Viçosa, Brazil.

8. Chagas, C.S., Fernandes Filho, E.I., Vieira, C.A.O., Schaefer, C.E.G.R., Carvalho Júnior, W. (2010). Topographic attributes and Landsat7 data in the digital soil mapping using neural networks. Pesq. agropec. bras. 45, 497–507. https://doi.org/10.1590/S0100-204X2010000500009.

9. Costa, J.L.P.O., Cavalcanti, A.P.B. (2010). Phytogeography of the delta plain of the Parnaíba River, Piauí/Maranhão - Brasil: analysis of the distribution of species and anthropogenic interference. Observatorium 2, 84–92. https://seer.ufu.br/index.php/Observatorium/article/view/44199

10. Dias, L.M., Coelho, R.M., Valladares, G.S., Assis, A.C.C., Ferreira, E.P., Silva, R.C. (2016). Soil class prediction by data mining in an area of the sedimentary São Francisco basin. Pesq. agropec. bras. 51, 1396–1404. https://doi.org/10.1590/s0100-204x2016000900038.

11. Embrapa (Brazilian Agricultural Research Corporation) (1979). *National Land Survey and Conservation Service (Rio de Janeiro, RJ)*. Summary of the 10th Technical Meeting of Soil Survey. Rio de Janeiro.

12. Gao, B.C. (1996). NDWI: A Normalized Difference Water Index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.

13. Guedes, H.A.S., Silva, D.D. (2012). Comparison between hydrographically conditioned digital elevation models in the morphometric charaterization of watersheds. Eng. Agríc. 32, 932–943. https://doi.org/10.1590/S0100-69162012000500012.

14. Hengl, T., Rossiter, D.G. (2003). Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Science Society of America Journal 67, 1810–1822. https://doi.org/10.2136/sssaj2003.1810.

15. Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schimdt, M.G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma* 265, 62–77. https://doi.org/10.1016/j.geoderma.2015.11.014.

16. Höfig, P., Giasson, E., Vendrame, P.R.S. (2014). Digital soil mapping based on map extrapolation between physiographically similar areas. *Pesq. Agropec. Bras.* 49, 958–66. https://doi.org/10.1590/S0100-204X2014001200006.

17. Horn, B.K.P. (1981). *Hill shading and the reflectance map.* Proceedings of the Institute of Electrical and Electronic Engineers 69, 14–47. https://doi.org/10.1109/PROC.1981.11918.

18. Jensen, J. R. (2009). *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*. Parêntese, São José dos Campos.

19. Kempen, B., Brus, D. J., Stoorvogel, J. J., Heuvelink, G. B. M., Vries, F. (2012). Efficiency Comparison of Conventional and Digital Soil Mapping for Updating Soil Maps. Soil Science Society of America Journal 76, 2097–2115. https://doi.org/10.2136/sssaj2011.0424.

20. Kheir, R.B., Greve, M. H., Abdallah, C., Dalgaard, T. (2010). Spatial soil zinc content distribution from terrain parameters: A GIS-based decision-tree model in Lebanon. Environmental Pollution 158, 520–528. https://doi.org/10.1016/j.envpol.2009.08.009.

21. Landis, J.R., Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics 33, 159–174. https://doi.org/10.2307/2529310.

22. Marengo, J.A., Torres, R.R., Alves, L.M. (2017). Drought in Northeast Brazil—past, present, and future. Theoretical and Applied Climatology 129, 1189–1200. https://doi.org/10.1007/s00704-016-1840-8.

23. McBratney, A. B., Odeh, I.O.A.; Bishop, T.F.A., Dunbar, M.S., Shatar, T.M. (2000). An overview of pedometric techniques for use in soil survey. Geoderma 97, 293–327. https://doi.org/10.1016/S0016-7061(00)00043-4.

24. Mendonça, F., Danni-Oliveira, I.M. (2007). *Climatologia: noções básicas e climas do Brasil*. Oficina de Textos, São Paulo.

25. Minasny, B., McBratney, A.B. (2015). Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. http://dx.doi.org/10.1016/j.geoderma.2015.07.017.

26. Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson G.A. (1993). Soil attribute prediction using terrain analysis. Soil Science Society American Journal 57, 443–452. https://doi.org/10.2136/sssaj1993.03615995005700020026x.

27. Nolasco-Carvalho, C.C., Franca-Rocha, W., Uchoa, J.M. (2009). Digital soil map: a methodological proposal using fuzzy inference. Revista Brasileira de Engenharia Agrícola e Ambiental 13, 46–55. https://doi.org/10.1590/S1415-43662009000100007.

28. Portela, M.G.T., Espindola, G.M., Valladares, G.S., Amorim, J.V.A., Frota, J.C.O. (2020). Vegetation biomass and carbon stocks in the Parnaíba River Delta, NE Brazil. Wetlands Ecology and Management 28, 607–622. https://doi.org/10.1007/s11273-020-09735-y.

29. Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., Waterloo, M.J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. Remote Sensing of Environment 112, 3469–3481. https://doi.org/10.1016/j.rse.2008.03.018.

30. Rizzo, R., Demattê, J.A.M., Lepsch, I.F., Gallo, B.C., Fongaro, C.T. (2016). Digital soil mapping at local scale using a multi-depth Vis–NIR spectral library and terrain attributes. Geoderma 274, 18–27. https://doi.org/10.1016/j.geoderma.2016.03.019.

31. Ruiz, L.F.C., Ten Caten, A., Dalmolin, R.S.D. (2011). ADtoSIG Conversion Software. Federal University of Santa Maria, Santa Maria, Brazil.

32. Santos-Filho, F.S., Almeida Junior, E.B., Soares, C.J.R.S., Zickel, C.S (2010). Faces of the Salt Marshes of Delta Parnaíba, Norteastern Brasil. Revista Brasileira de Geografia Física 3, 218–227. https://doi.org/10.26848/rbgf.v3i3.232605.

33. Silveira, C.T., Oka-Fiori, C., Santos, L.J.C., Sirtoli, A.E., Silva, C.R., Botelho, M.F. (2013). Soil prediction using artificial neural networks and topographic attributes. Geoderma 195–196, 165–172. https://doi.org/10.1016/j.geoderma.2012.11.016

34. Sirtoli, A.E., Silveira, C.T., Montovani, L.E., Sirtoli, A.R.A., Oka-Fiori, C. (2008). Landscape atributes derived from digital elevation model and relationships with soils. Scientia Agraria 9, 317–329. http://dx.doi.org/10.5380/rsa.v9i3.11517.

35. Sobrinho, T.A., Oliveira, P.T.S., Rodrigues, D.B.B., Ayres, F.M. (2010). Automatic delimitation of watershed using data SRTM. Eng. Agríc. 30, 46–57. https://doi.org/10.1590/S0100-69162010000100005.

36. Sousa, R.S. (2015). *Planície Costeira do Estado do Piauí: mapeamento das unidades de paisagem, uso e cobertura da terra e vulnerabilidade ambiental.* Master's degree dissertation, Federal University of Piauí, Teresina, Brazil.

37. Taghizadeh-Mehrjardia, R., Nabiollahi, K., Kerryc, R. (2016). Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* 266, 98–110. https://doi.org/10.1016/j.geoderma.2015.12.003.

38. Ten Caten, A. (2011). *Digital soil mapping: Methods to meet the demand for soil spatial information.* Ph.D. Thesis, Federal University of Santa Maria, Santa Maria, Brazil. https://repositorio.ufsm.br/handle/1/3326?show=full Acessed 12 October 2020.

39. Valeriano, M.M., (2008). Dados topográficos, in: Florenzano, T.G. (Org.). *Geomorfologia, conceitos e tecnologias atuais.* Oficina de Textos, São Paulo, pp.72–104.

40. Valladares, G.S., Cabral, L.J.R.S. (2017). Mapeamento geológico da Planície do Delta do Parnaíba-PI. Symposium of Northeastern Geology, 27, 2017, João Pessoa. SGNE, *Anais...* João Pessoa. http://www.geologiadonordeste.com.br/anais/index_sessao.php#resumos Acessed 10 October 2020.

41. Witten, I.H., Frank E., Hall, M.A. (2011). *Data mining: Practical machine learning tools and techniques*, third ed. Morgan Kaufmann, San Francisco.

42. Witten, I.H., Frank, E. (2005). *Data mining: practical machine learning tools and techniques*, second ed. Morgan Kaufmann, San Francisco.

# Tables

Table 4 – Confusion matrix of data set 2 (field samples)

|  | GLeu | GLti | AReu | SNap | MD | SB | PA |
|---|---|---|---|---|---|---|---|
| GLeu | 2 | 0 | 1 | 1 | 0 | 0 | **4** |
| GLti | 0 | 2 | 0 | 0 | 0 | 0 | **2** |
| AReu | 1 | 0 | 1 | 1 | 0 | 1 | **4** |
| SNap | 1 | 0 | 0 | 4 | 0 | 0 | **5** |
| MD | 0 | 0 | 0 | 0 | 1 | 1 | **2** |
| SB | 0 | 0 | 0 | 0 | 1 | 3 | **4** |
| UA | **4** | **2** | **2** | **6** | **2** | **5** | **21** |

Legend: UA = user's accuracy; PA = producer's accuracy.

Table 5 – Confusion matrix of data set 2 (field samples)

|  | GL | AR | FL | MD | SB | PA |
|---|---|---|---|---|---|---|
| GL | 3 | 0 | 0 | 0 | 0 | 3 |
| AR | 0 | 2 | 3 | 0 | 0 | 5 |
| FL | 0 | 1 | 6 | 0 | 0 | 7 |
| MD | 0 | 1 | 0 | 1 | 0 | 2 |
| SB | 0 | 1 | 0 | 1 | 2 | 4 |
| UA | 3 | 5 | 9 | 2 | 2 | **21** |

Legend: UA = user's accuracy; PA = producer's accuracy.

Table 6 – Confusion matrix of the comparison between mapping 1 and the traditional mapping

|  |  | GLeu | GLti | AReu | SNap | MD+SB | PA |
|---|---|---|---|---|---|---|---|
| PRE-EXISTING MAP Cabral (2018) | GLeu | 12.13 | 1.66 | 7.12 | 7.27 | 0.61 | 28.79 |
|  | GLti1+GLti2 | 12.13 | 1.66 | 7.12 | 7.27 | 0.51 | 28.69 |
|  | AReu1 | 5.83 | 0.36 | 57.76 | 5.27 | 10.35 | 79.57 |
|  | SNap | 9.63 | 0.08 | 15.23 | 26.27 | 3.63 | 54.84 |
|  | AReu2 | 0.52 | 0.72 | 5.28 | 0.34 | 24.45 | 31.32 |
| UA |  | 40.24 | 4.47 | 92.52 | 46.42 | 39.55 | **223.20** |

Legend: UA = user's accuracy; PA = producer's accuracy.

Table 7 – Confusion matrix of the comparison between mapping 2 and the traditional mapping

|  |  | GL | AR | FL | MD+SB | PA |
|---|---|---|---|---|---|---|
| PRE-EXISTING MAP Cabral (2018) | GLti1+GLti2+GLeu | 37.16 | 16.64 | 7.43 | 0.47 | 61.70 |
|  | AReu1 | 0.77 | 64.93 | 9.18 | 4.69 | 79.57 |
|  | SNap | 0.80 | 14.26 | 38.91 | 0.88 | 54.84 |
|  | AReu2 | 0.83 | 5.65 | 0.63 | 24.21 | 31.32 |
| UA |  | 39.55 | 101.47 | 56.14 | 30.25 | **227.42** |

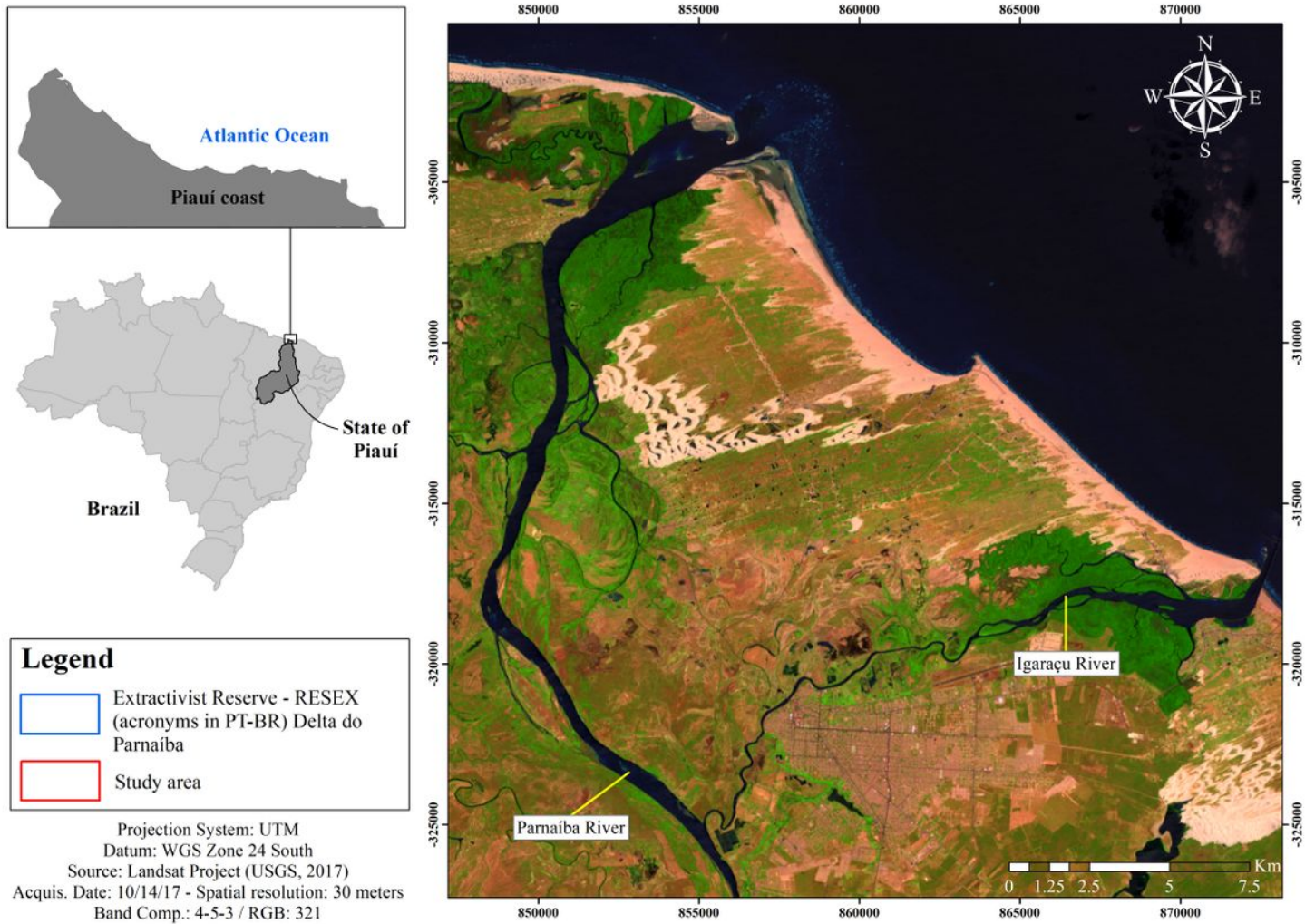Legend: UA = user's accuracy; PA = producer's accuracy.

# Figures

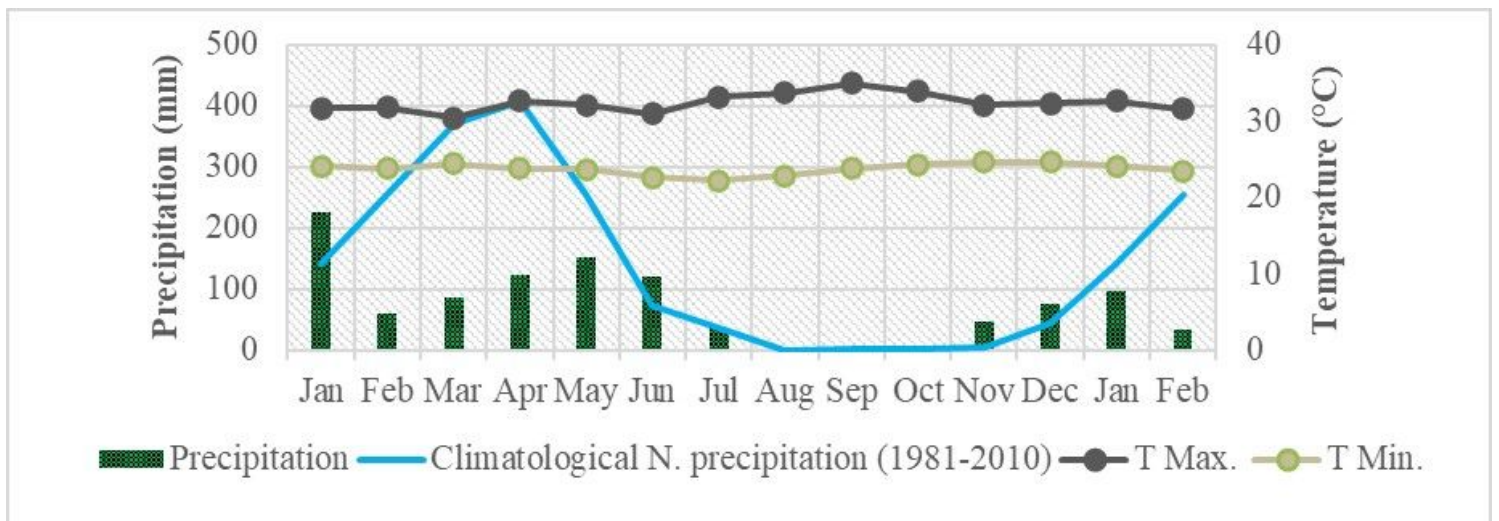**Figure 1**

Location of the study area

## Figure 2

Data of precipitation (mm) and maximum and minimum temperatures (°C) of the study area from January 2016 to February 2017 and climatological normal precipitation from 1981 to 2010. Source: INMET (Brazil, 2018)
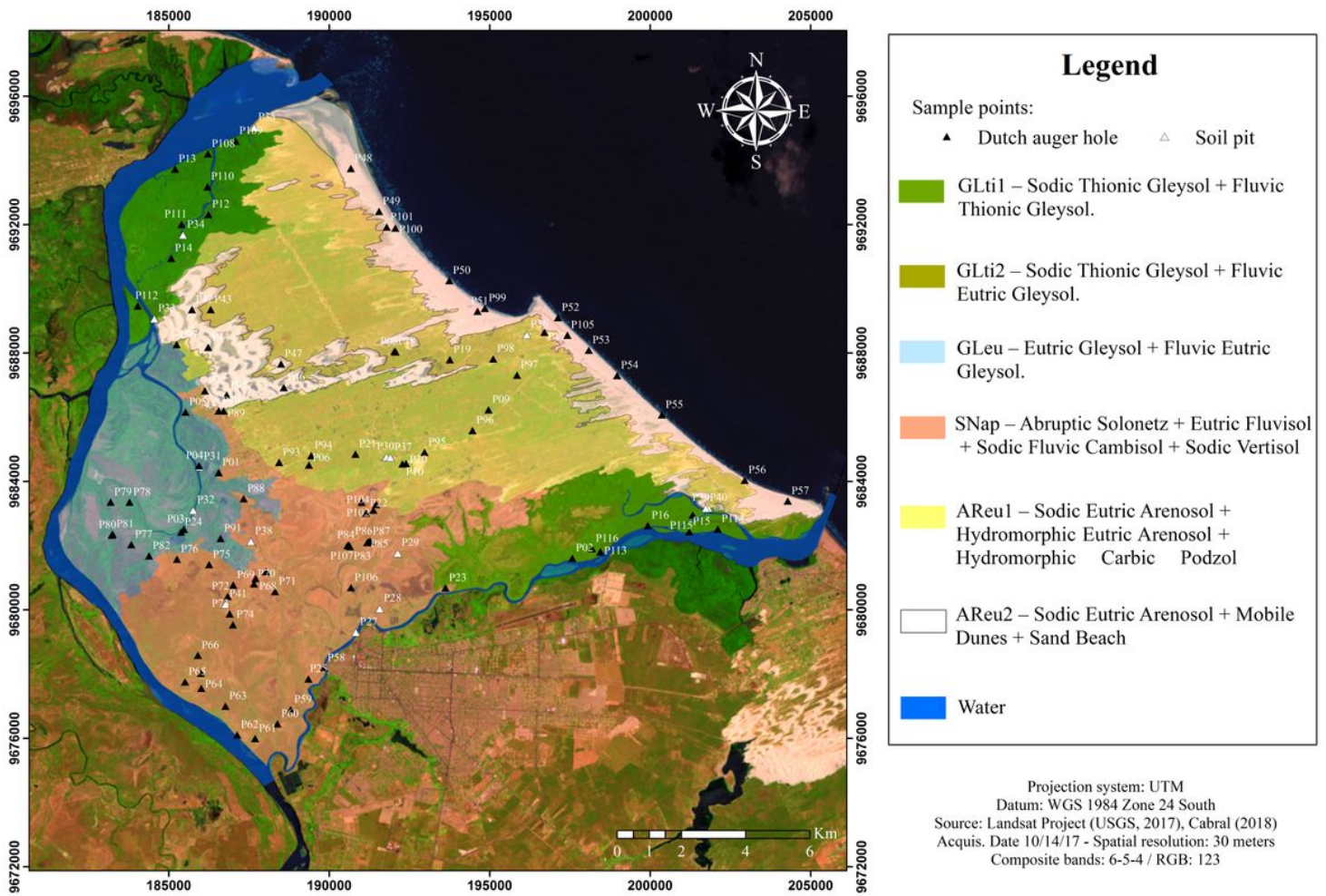


## Figure 3

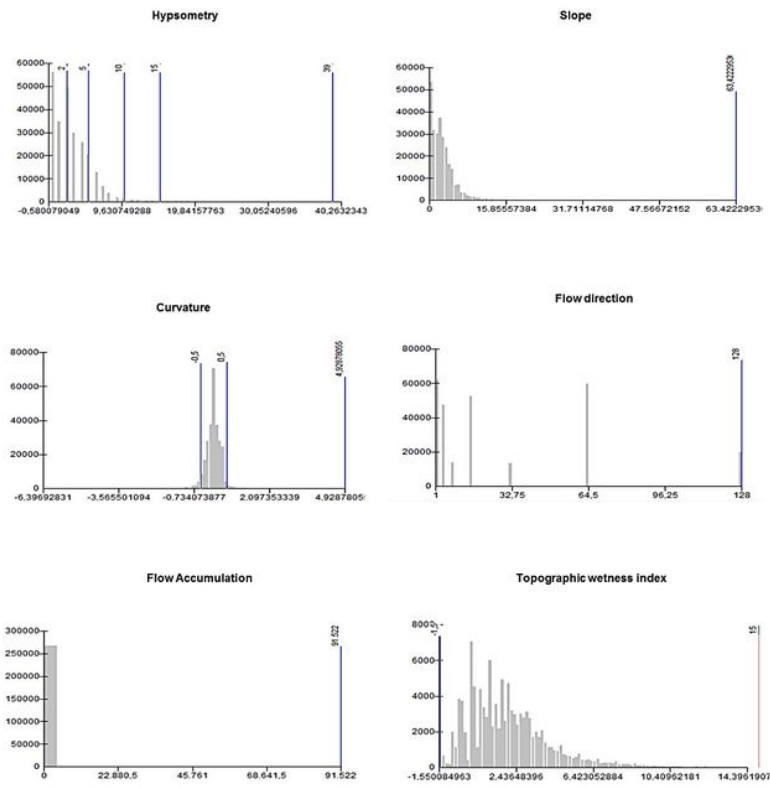Location of soil sampling points

Source: Adapted from Cabral (2018).

**Figure 4**

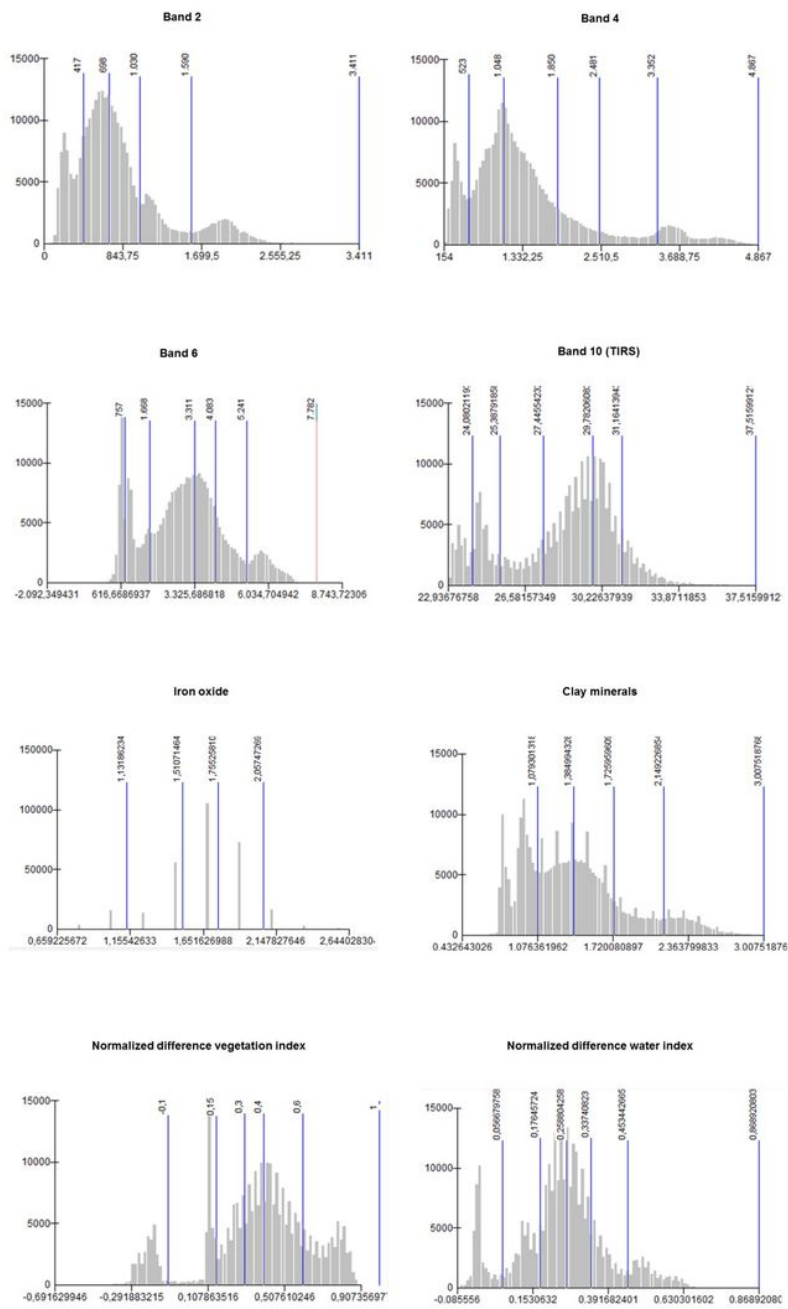Frequency histogram of the geomorphometric attributes.

**Figure 5**

Frequency histogram of the spectral attributes.

```
=== Run information ===

Scheme:        weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:      solos
Instances:     98
Attributes:    15
               B2
               B4
               B6
               B10
               IRON
               CLAY
               NDVI
               NDWI
               ALT
               DEC
               CUR
               FD
               FAC
               TWI
               class
Test mode:     10-fold cross-validation
```

**Matrix 1**

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

NDVI <= 0.151539
|   IRON <= 1.8867: SB (11.0/1.0)
|   IRON > 1.8867: MD (4.0)
NDVI > 0.151539
|   B10 <= 28398
|   |   B10 <= 27662: GLti (15.0)
|   |   B10 > 27662
|   |   |   TWI <= 0.108637: GLeu (3.0)
|   |   |   TWI > 0.108637: GLti (3.0/1.0)
|   B10 > 28398
|   |   IRON <= 1.64329: AReu (6.0)
|   |   IRON > 1.64329
|   |   |   ALT <= 5
|   |   |   |   B10 <= 30316
|   |   |   |   |   B2 <= 501
|   |   |   |   |   |   FAC <= 3
|   |   |   |   |   |   |   B2 <= 417: GLeu (5.0)
|   |   |   |   |   |   |   B2 > 417: SNap (5.0/1.0)
|   |   |   |   |   |   FAC > 3: SNap (6.0)
|   |   |   |   |   B2 > 501
|   |   |   |   |   |   B10 <= 30063
|   |   |   |   |   |   |   TWI <= 4.29462: AReu (12.0/1.0)
|   |   |   |   |   |   |   TWI > 4.29462: SNap (2.0)
|   |   |   |   |   |   B10 > 30063: GLeu (6.0/1.0)
|   |   |   |   B10 > 30316: SNap (15.0/1.0)
|   |   |   ALT > 5
|   |   |   |   IRON <= 2.11818: MD (3.0/1.0)
|   |   |   |   IRON > 2.11818: SNap (2.0)

Number of Leaves  :     15

Size of the tree  :     29
```

```
=== Confusion Matrix ===

 a  b  c  d  e  f   <-- classified as
 8  0  2  6  0  0 |  a = GLeu
 1 14  1  0  1  0 |  b = GLti
 4  0  6  4  3  2 |  c = AReu
 6  1  5 17  1  0 |  d = SNap
 0  0  1  1  2  2 |  e = MD
 0  0  1  0  1  8 |  f = SB
```

**Matrix 2**

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

NDVI <= 0.151539
|   IRON <= 1.8867: SB (11.0/1.0)
|   IRON > 1.8867: MD (4.0)
NDVI > 0.151539
|   B10 <= 28398: GL (21.0/1.0)
|   B10 > 28398
|   |   B10 <= 30063
|   |   |   B2 <= 490
|   |   |   |   CLAY <= 1.97182
|   |   |   |   |   B4 <= 797
|   |   |   |   |   |   B10 <= 29188: GL (4.0)
|   |   |   |   |   |   B10 > 29188: FL (3.0)
|   |   |   |   |   B4 > 797: FL (9.0)
|   |   |   |   CLAY > 1.97182: AR (2.0)
|   |   |   B2 > 490
|   |   |   |   ALT <= 5
|   |   |   |   |   TWI <= 4.29462: AR (16.0/1.0)
|   |   |   |   |   TWI > 4.29462: FL (2.0)
|   |   |   |   ALT > 5
|   |   |   |   |   B2 <= 600: AR (2.0/1.0)
|   |   |   |   |   B2 > 600: MD (2.0)
|   |   B10 > 30063: FL (22.0)

Number of Leaves  :     12

Size of the tree  :     23
```

```
=== Confusion Matrix ===

 a  b  c  d  e   <-- classified as
18  2  3  1  0 |  a = GL
 2 11  3  1  2 |  b = AR
 3  3 33  0  0 |  c = FL
 0  2  0  2  2 |  d = MD
 0  0  0  0 10 |  e = SB
```

## Figure 6

Result of processing of the data matrices in WEKA software (generated classification models and confusion matrices)

**Figure 7**

Digital soil mapping generated from the data matrix 1

**Figure 8**

Digital soil mapping generated from the data matrix 2

**Figure 9**

Analysis of agreement between the pre-existing soil map and the digital mappings.