

Division - Soil in Space and Time | Commission - Pedometrics

Digital soil mapping for the Parnaíba River delta, Brazilian semiarid region

João Victor Alves Amorim^{(1)*} , Gustavo Souza Valladares⁽²⁾ , Marcos Gervasio Pereira⁽³⁾ , Mirya Grazielle Torres Portela⁽⁴⁾  and Andréa Maciel Lima⁽⁵⁾ 

⁽¹⁾ Universidade Federal do Piauí, Colégio Técnico de Bom Jesus, Campus Professora Cinobelina Elvas, Bom Jesus, Piauí, Brasil.

⁽²⁾ Universidade Federal do Piauí, Campus Ministro Petrônio Portella, Teresina, Piauí, Brasil.

⁽³⁾ Universidade Federal Rural do Rio de Janeiro, Instituto de Agronomia, Campus Seropédica, Seropédica, Rio de Janeiro, Brasil.

⁽⁴⁾ Universidade Federal do Piauí, Colégio Técnico de Floriano, Campus Amílcar Ferreira Sobral, Floriano, Piauí, Brasil.

⁽⁵⁾ Universidade Federal de Goiás, Instituto de Estudos Socioambientais, Campus Samambaia, Goiânia, Goiás, Brasil.

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 two data matrices: one elaborated for the mapping units of the pre-existing conventional pedological map and the other for a set of associations determined based on the characteristics of the landscape of the study area with close correlation with the existing soils, mainly due to the source material. 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 using a pre-existing traditional soil map of the area. The results revealed that the attributes derived from satellite images stood out from those derived from DEM in predicting soil groups. Based on the validation coefficients applied (overall accuracy, Kappa index, user's accuracy and producer's accuracy), the classification quality was satisfactory, despite the complexity of the environment.

Keywords: soil prediction, decision trees, pedometry, soil survey, remote sensing, morphometry.



* **Corresponding author:**

E-mail: amorim@ufpi.edu.br

Received: October 03, 2022

Approved: February 27, 2023

How to cite: Amorim JVA, Valladares GS, Pereira MG, Portela MGT, Lima AM. Digital soil mapping for the Parnaíba River delta, Brazilian semiarid region. Rev Bras Cienc Solo. 2023;47:e0220160. <https://doi.org/10.36783/18069657rbc20220160>

Editors: José Miguel Reichert  and Adelson Paulo de Araújo 

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.



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 and improve conventional work of pedology, enabling larger areas to be mapped (Nolasco-Carvalho et al., 2009).

Accurate and actual maps of soil properties are essential for evaluating multiple aspects of agricultural land management, including soil production potential, water retention, soil erosion and total carbon stock (Žížala et al., 2022). The application of pedometry techniques depends on the objective and 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 has 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 (Chagas et al., 2010; Silveira et al., 2013; Brungard et al., 2015; Heung et al., 2016), in addition to models that apply decision trees (DT) (Höfig et al., 2014).

The DT approach has been used, among several reasons, for having the advantages of enabling the explicit expression of soil landscape relationships, grouping and search for patterns of distribution of soil classes, besides enabling the understanding of how these data are interrelated (Kheir et al., 2010). It is important to provide solid and rational criteria that justify the continuity of funding for pedometric research in Brazil, considering that soil information guides plans in important decisions, such as the appropriation of areas of high environmental fragility, such as coastal plains, for example (Rosa, 2012).

Coastal plains are characterized as environments of great geological, geomorphological, pedological and landscape diversity. However, this fragile environment has high rates of urban growth and demographic concentration, which characterizes it as an area of physical and social vulnerability (Barros and Muehe, 2010; Marandola Júnior et al., 2013). The coastal zone of the state of Piauí, Brazil, presents landscapes composed of beaches, fluvio-marine plains, fluvio-lacustrine plains, dunes, wind deflation plains, in addition to deltaic islands, subjected to human interventions that alter the natural dynamics of these environmental systems.

Among these landscapes is located the Parnaíba River Delta, an environment made up of several fragile ecosystems and poorly consolidated materials, with a high scenic beauty, where several erosive and deposition processes act on it, characterizing, in this way, a highly dynamic environment. For the conservation of this natural environment and sustainability in the exploitation of natural resources, studies are needed to investigate the soils, their limitations and potential, since these areas, as well as other coastal environments in tropical regions of the world, are important centers of environmental maintenance and socioeconomic of societies. 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.

MATERIALS AND METHODS

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 Igarapé 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 km² and delimited by the coordinates 848137/-301083 and 873137/-327576.

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 No. 9,985 of July 18, 2000, which establishes the Brazilian National System of Nature Conservation Units (SNUC-acronym in Portuguese) (Brasil, 2000). It classifies conservation units into full protection and sustainable use units, and the two aforementioned units are classified as sustainable use units (Figure 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

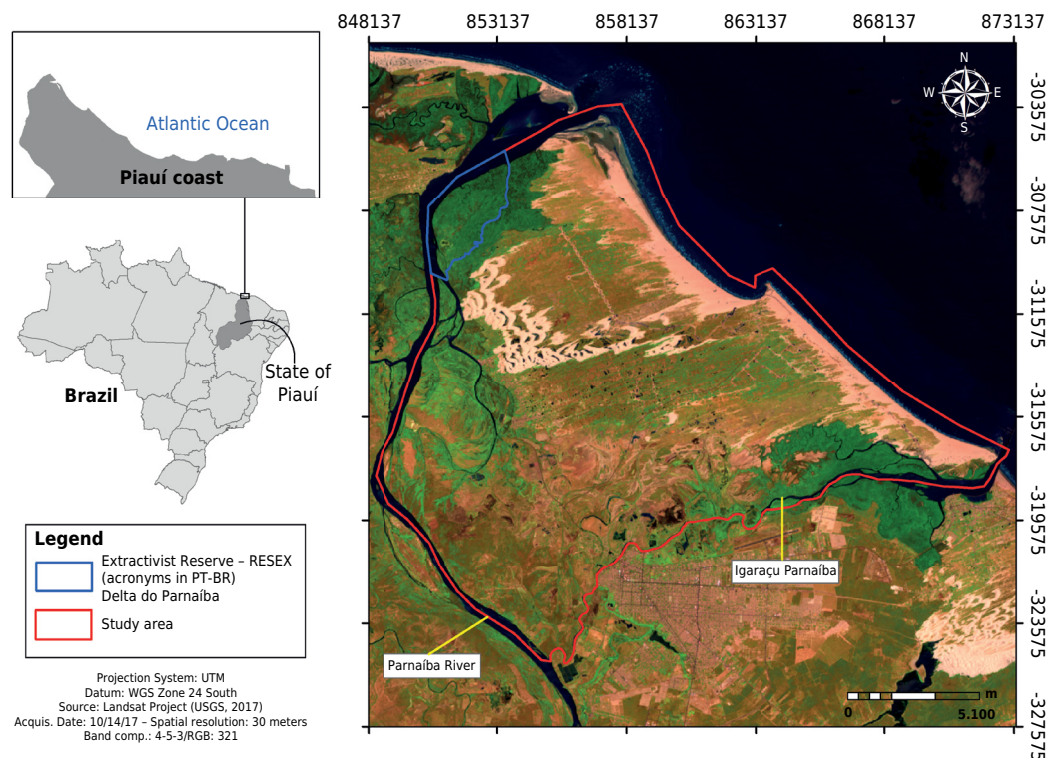


Figure 1. Location of the study area.

average temperature reaching 35 °C in September. Compared to the climatological normal of precipitation in 30 years (Brasil, 2018), 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 never seen in previous decades (Marengo et al., 2017).

The geology is formed by 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 the predominance of depositional processes, both of the continental type, as is the case of the fluvial plain, and 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.

As a conservation unit of sustainable use, classified by the SNUC, the study area is occupied by urban spaces, agricultural activities and wind energy production, with restricted access areas. The region has 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 in the area is constituted mostly of woody species, characterized by trees spaced with irregular crowns, 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).

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 (pixels with 90 m of size) of the area and extraction of vector planialtimetric data from the topographic sheet of Geographic Service Directorate of the Brazilian Army (Brasil, 1972) Parnaíba sheet (SA 24 Y-A-IV) at a scale of 1:100,000. In geographic information system (GIS), the variables elevation, slope, curvature, flow direction, flow accumulation and topographic wetness index were generated at a scale of 1:100,000.

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 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 and the visualization of the relief (Rennó et al., 2008).

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

The topographic wetness index is used to characterize the spatial distribution of surface saturation zones and water content in landscapes (Sirtoli et al., 2008). This index further demonstrates the effects of relief on the location and extension of water accumulation areas (Moore et al., 1993). 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 (Silveira et al., 2013). It is defined as a function of slope and the contributing area per unit width orthogonal to the flow direction (Chagas, 2006) obtained through equation 1:

$$TWI = \ln \left(\frac{As}{\tan\beta} \right) \quad \text{Eq. 1}$$

in which: As is the contributing area multiplied by the size of the grid cell by m^2 ; and β 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 a spatial resolution of 30 m, subjected to orthorectification and radiometric corrections and are under the coordinate system UTM Dtuna WGS 1984, zone 24 South.

Band 2 is used more effectively in bathymetric mappings, distinguishing soil from vegetation and deciduous vegetation from tree vegetation (Barsi et al., 2014). 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 the 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 a thermal band, in the infrared range, with wavelength from 10.6 to 11.2 μm , 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 m, but they are resampled to 30 m 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:

$$\text{CLAY} = \frac{\text{BAND 6}}{\text{BAND 7}} \quad \text{Eq. 2}$$

$$\text{IRON} = \frac{\text{BAND 4}}{\text{BAND 2}} \quad \text{Eq. 3}$$

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 equation 4:

$$\text{NDVI} = \frac{\text{BAND 6} - \text{BAND 4}}{\text{BAND 6} + \text{BAND 4}} \quad \text{Eq. 4}$$

in which BAND 6 is the near infrared; and BAND 4 is red.

The NDWI was designed to delimitate open water environments, automating the determination of the threshold between water and land (terrestrial vegetation and soils). In this perspective, the equation proposed by Gao (1996) was applied to highlight the water bodies, using reflectance data of the bands corresponding to 860 and 1240 nm, respectively (Equation 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:

$$\text{NDWI} = \frac{\text{BAND 5} - \text{BAND 6}}{\text{BAND 5} + \text{BAND 6}} \quad \text{Eq. 5}$$

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

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 figure 2. The samples were stratified and drawn within each mapping unit established in the conventional soil map (Cabral, 2018) (Figure 2). The division was carried out 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. 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).

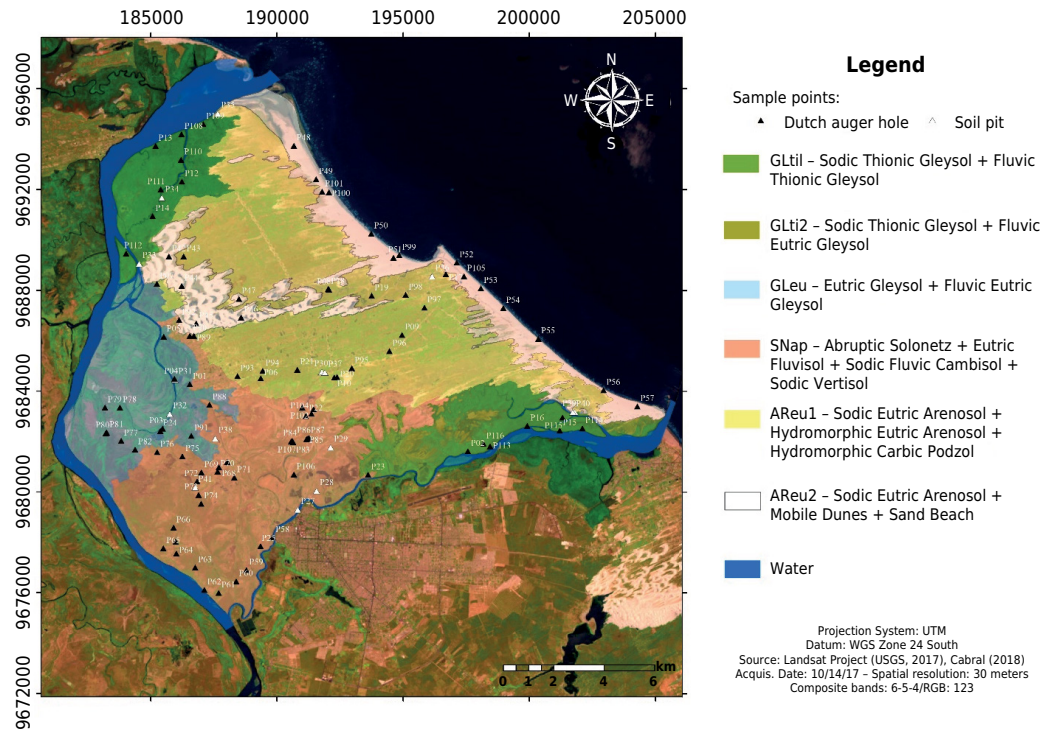


Figure 2. Location of soil sampling points. Source: Adapted from Cabral (2018).

Table 1. Description of data matrices created for DSM

Matrices	Predictive attributes	Classes
1) Map units adapted from Cabral (2018)	All parameters: elevation, 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

TWI: topographic wetness index; IRON: iron oxide index; CLAY: clay minerals index. Classification according to the World Reference Base for Soil Resources/Food and Agriculture Organization (WRB/FAO) and Brazilian Soil Classification System (SIBCS), in parentheses: GLeu – Eutric Gleysol (*Gleissolo Háptico*), SCgl – Gleyic Solonchack (*Gleissolo Sáfico*), GLum – Umbric Gleysol (*Gleissolo Melânico*), GLti – Thionic Gleysol (*Gleissolo Tiomórfico*), AReu – Eutric Arenosol (*Neossolo Quartzarênico*), PZcb – Carbic Podzol (*Espodossolo Humilúvico*), SNap – Abruptic Solonetz (*Planossolo Nátrico*), FLeu – Eutric Fluvisol (*Neossolo Flúvico*), CMfv – Fluvic Cambisol (*Cambissolo Flúvico*), VRso – Sodic Vertisol (*Vertissolo Háptico*).

In Weka, for analysis of the datasets, the model was adjusted and executed. The performance of the J48 classifier algorithm, corresponding to the DT method, was verified. The result of processing by the algorithm, which automatically generates an accuracy assessment, was used for mapping. 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). This technique has become a standard for data mining applications (Witten et al., 2011). Furthermore, according to studies evaluated by these authors, number 10 (ten) should be adopted as the standard value for the number of data partitions (K).

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 matrixes with the best performance in classification was exported to a 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 maps were 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. All maps were produced at a scale of 1:100,000.

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 (Figure 2). These data were compared with the maps predicted through 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 classification accuracy, according to Landis and Koch (1977) (Table 2). For comparison with the pre-existing soil map, validation was performed based on all pixels.

RESULTS AND DISCUSSION

The DEM generated six geomorphometric attributes for the study area: elevation, slope, curvature, flow direction, flow accumulation and topographic wetness index. For elevation, the values varied from 0 to 5 m above sea level, and the ranges from 10 to 15 and >15 m 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 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 a 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

Table 2. Image quality classification according to kappa index ranges

Kappa coefficient	Classification Accuracy
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

Source: Adapted from Landis and Koch (1977).

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). The TWI will display negative values when the catchment area is lower than $\tan(B)$, since the natural logarithm of any value lower than one 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 3 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 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 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

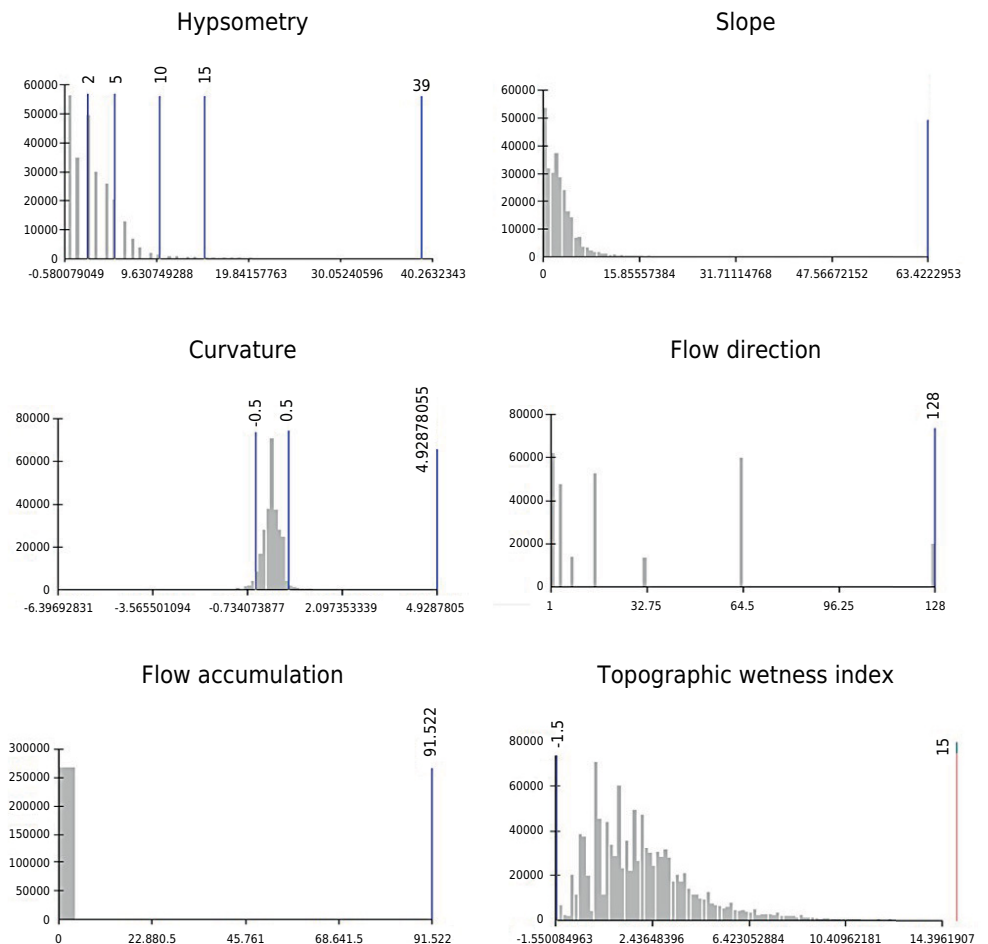


Figure 3. Frequency histogram of the geomorphometric attributes.

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. They 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: *Gleissolo Tiomórfico* (Thionic Gleysol) + *Gleissolo Melânico* (Umbric Gleysol) + *Gleissolo Sáfico* (Gleyic Solonchack) + *Organossolo Tiomórfico* (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: *Planossolo Nátrico* (Abruptic Solonetz) + *Cambissolo Flúvico* (Fluvic Cambisol) + *Neossolo Flúvico* (Eutric Fluvisol) + *Vertissolo Háplico* (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 primary minerals, mainly sands constituted by quartz found in mobile dunes, sandspits and beach areas.

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 dunes and paleo-dunes. In these places, herbaceous and/or shrubby plant species are predominant, 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 *Neossolos* (Arenosols) can occur and with substantial restriction to drainage, that is, the presence of hydromorphism. This situation of relief causes the 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 the presence of herbaceous and/or shrubby undergrowth and, associated with it, mainly *Neossolos Quartzarênicos* (Eutric Aerosol). The values between 0.3 and 0.4 classified the areas of open natural fields with *Neossolo Flúvico* (Eutric Fluvisol). Values from 0.4 to 0.6 indicate areas with a predominance of more shrub/tree sized vegetation, where it is possible to find associations of *Planossolo Nátrico* (Abruptic Solonetz) + *Cambissolo Flúvico* (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 the occurrence of hydromorphic soils.

The 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 *Planossolo Nátrico* (Abruptic Solonetz) and *Cambissolo Flúvico* (Fluvic Cambisol) occur. Between 0.34 and 0.45, it was possible to classify the wind plain, indicating the

occurrence of sandy soils: *Neossolo Quartzarênico* (Eutric Arenosol) and *Espodossolo Humilúvico* (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 4 shows the frequency histogram of the spectral attributes.

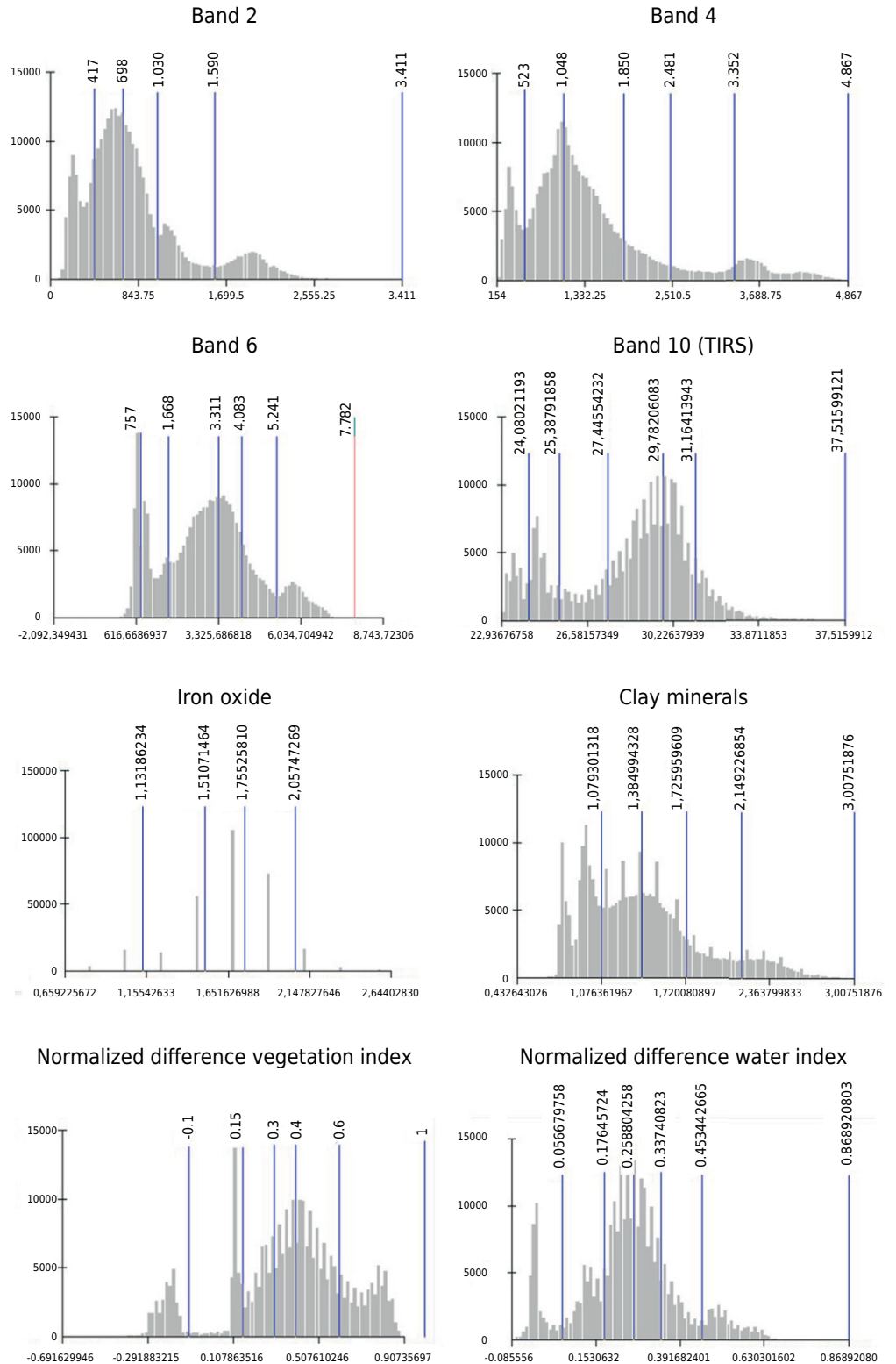


Figure 4. Frequency histogram of the spectral attributes.

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

CRI: Correctly Ranked Instances.

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 Weka program performed the analyses, 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.

For matrix 1, the classifier model selected the parameters NDVI, IRON, thermal band, band 2, elevation and flow accumulation as the ones with the greatest predominance for the classification. Based on the confusion matrix, it is observed that of the seventeen cells previously indicated as Thionic Gleysol (GLti), fourteen were correctly classified, with a hit percentage of 82.3 %. Only half of the cells indicated as Eutric Gleysol (GLEu) were correctly classified, and of the total, 37.5 % were classified as Abruptic Solonetz (SNap). A justification is that the Gleyic Solonchack (SCgl) present are located in environments similar to those of Eutric Fluvisols (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 (Figure 6).

Based on the classifier model of matrix 2 (Figure 5), it is verified that the parameters used by the algorithm were: NDVI, IRON, thermal band (band 10), band 2, CLAY, band 4, elevation and curvature. The NDVI was decisive for constructing the model, being able to individually separate two classes, SB (sand beach) and MD (mobile dunes), with the help of IRON. Another attribute of great representativeness was the thermal band, which identified the soils with gleyc characteristics (GL class) due to its ability to determine soil moisture.

The other parameters were considered in the classification to determine the other classes. However, it is worth mentioning that the geomorphometric attributes had lower relevance, and only elevation and curvature were entered into 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 a higher level of detailing is necessary to verify a greater correlation of morphometric attributes with soils.

The digital soil maps of matrices 1 and 2 were generated. Figure 6 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 *Neossolo Quartzarênico* (Eutric Arenosols) stand out, with simple grain structure. They contain the mineral quartz as predominant in all of their fractions. This class also comprise *Espodossolo Humilúvico* (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 *Planossolo Nátrico* (Abruptic Solonetz) stands out; these are soils with high sodium saturation, with prismatic or columnar structure. Its high textural gradient causes great

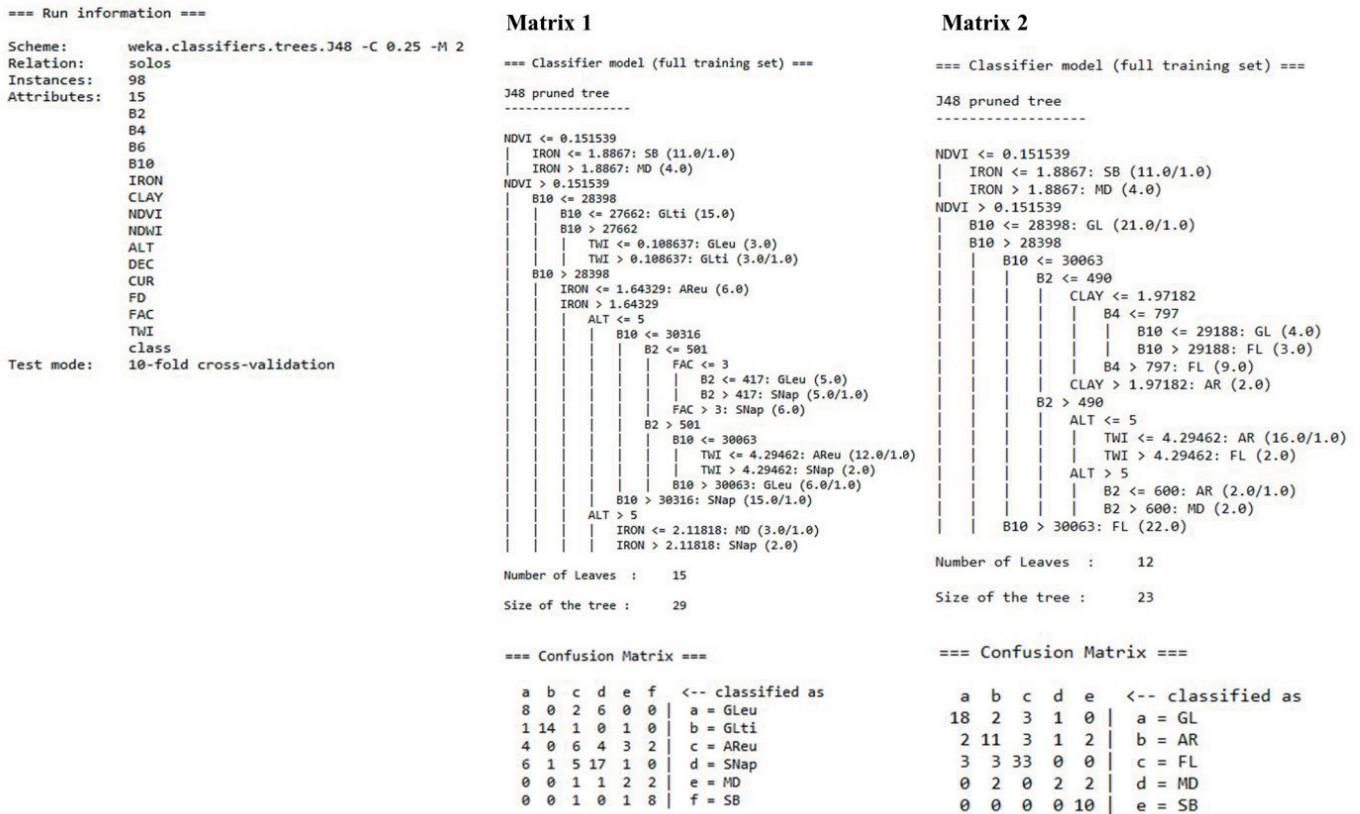


Figure 5. Result of processing the data matrices in WEKA software (generated classification models and confusion matrices). B2: Band 2 of the Landsat image; B4: Band 4 of the Landsat image; B6: Band 6 of the Landsat image; B10: Band 10 of the Landsat image; IRON: Iron oxide index; CLAY: Clay minerals index; NDVI: Normalized difference vegetation index; NDWI: Normalized difference water index; ALT: Elevation; DEC: Slope; CUR: Curvature; FD: Flow direction; FAC: Flow accumulation; TWI: Topographic wetness index; GLeu: Eutric Gleysol (*Gleissolo Háplico*); GLti: Thionic Gleysol (*Gleissolo Tiomórfico*); AReu: Eutric Arenosol (*Neossolo Quartzarênico*); SNap: Abruptic Solonetz (*Planossolo Nátrico*); GL: Soils with gleyic characteristics; AR: Soils with sand texture; FL: Soils with fluvic materials; MD: Mobile dunes; SB: Sand beach.

susceptibility to erosion, also favored by the low permeability of B horizon. There is also the occurrence of *Neossolo Flúvico* (Eutric Fluvisols), which are poorly evolved mineral soils formed from recent alluvial deposits. In addition to these, there are also *Cambissolo Flúvico* (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 *Vertissolo Háplico* (Sodic Vertisol), which consists of mineral material, with vertic characteristics in the horizons, the 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 if the chroma is less than or equal to 2.

Classes MD and SB extended for 27.59 km² (11.91 %) and 37.99 km² (16.41 %), respectively. Mobile dunes are large, individual-moving masses of sand consisting of simple and/or composite wind dunes and large strips of sand stretched to 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 through an error matrix, expressed in table 4 showed an overall accuracy of 61.90 %. The same is true for kappa, which for this map reached 0.52, a good agreement according to criteria of Landis and Koch (1977).

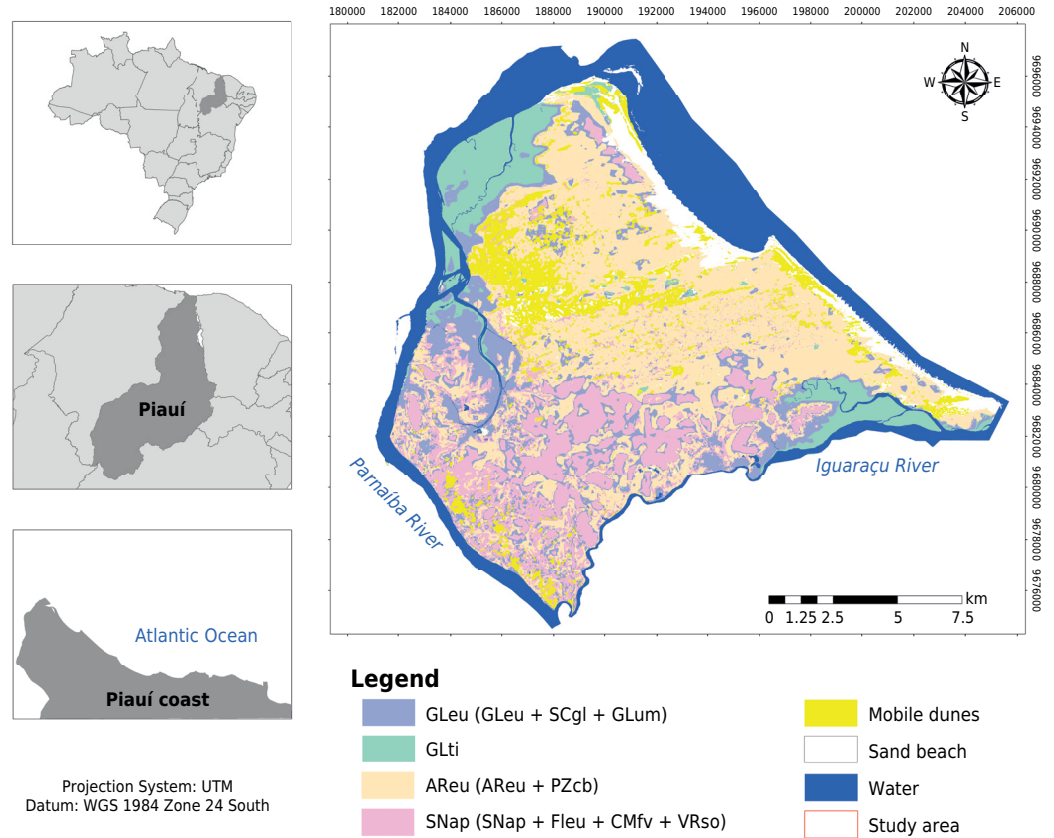


Figure 6. Digital soil mapping generated from the data matrix 1. GLeu: Eutric Gleysol (*Gleissolo Háplico*); SCgl: Gleyic Solonchack (*Gleissolo Sálico*); GLum: Umbric Gleysol (*Gleissolo Melânico*); GLti: Thionic Gleysol (*Gleissolo Tiomórfico*); AREu: Eutric Arenosol (*Neossolo Quartzarênico*); PZcb: Carbic Podzol (*Espodossolo Humilúvico*); SNAp: Abruptic Solonetz (*Planossolo Nátrico*); FLeu: Eutric Fluvisol (*Neossolo Flúvico*); CMfv: Fluvic Cambisol (*Cambissolo Flúvico*); VRso: Sodic Vertisol (*Vertissolo Háplico*).

Table 4. Confusion matrix of data set 1 (field samples)

Soil mapping unit classes	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

GLeu: Eutric Gleysol (*Gleissolo Háplico*); GLti: Thionic Gleysol (*Gleissolo Tiomórfico*); AREu: Eutric Arenosol (*Neossolo Quartzarênico*); Snap: Abruptic Solonetz (*Planossolo Nátrico*); MD: Mobile dunes; SB: Sand beach; UA: User's accuracy; PA: Producer's accuracy.

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 figure 7. Based on the image, it was verified that the most comprehensive class for the study area was sand texture

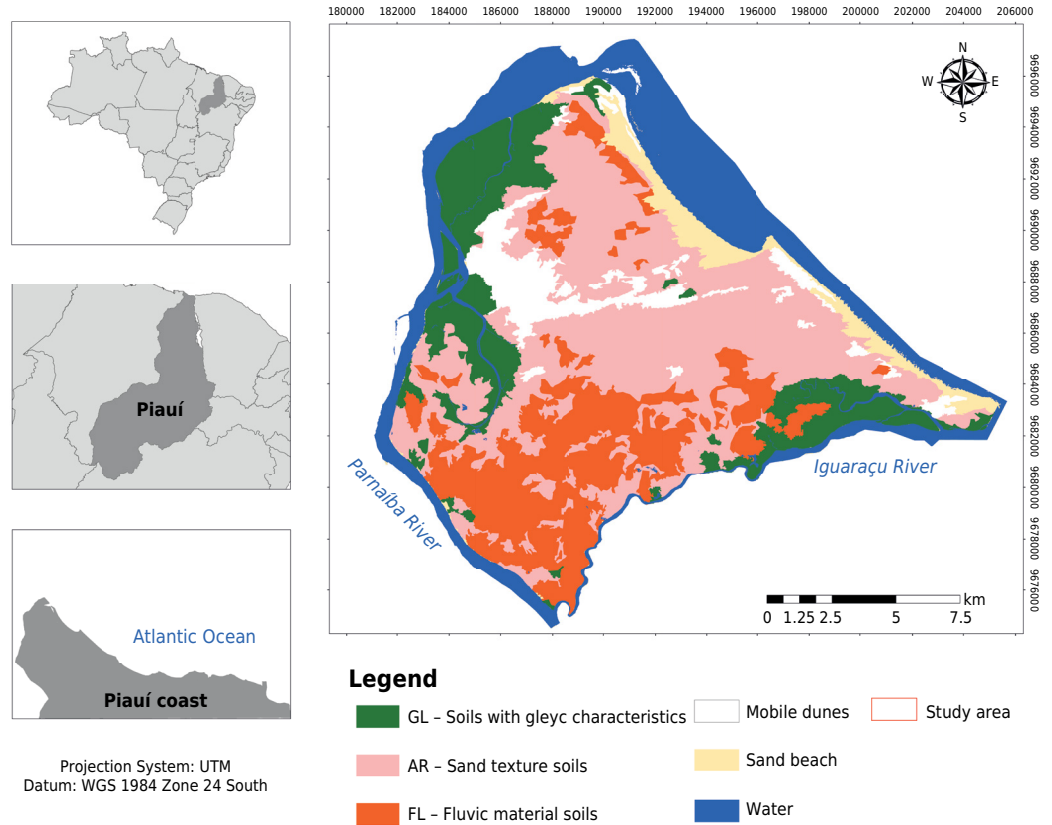


Figure 7. Digital soil mapping generated from the data matrix 2.

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

Soil grouping classes	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

GL: Soils with gleyc characteristics; AR: Soils with sand texture; FL: Soils with fluvic materials; MD: Mobile dunes; SB: Sand beach; UA: User's accuracy; PA: Producer's accuracy.

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 Igarapé River, and the Atlantic Ocean. Classes MD and SB extended for 22.99 km² (9.9 %) and 13 km² (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 %. 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.

Validation was also performed from the pre-existing traditional map of Cabral (2018) (Figure 2). 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 figure 8, where it is possible to verify the areas of agreement and disagreement between the two mappings.

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

GLEu: Eutric Gleysol (*Gleissolo Háplico*); GLti: Thionic Gleysol (*Gleissolo Tiomórfico*); AREu: Eutric Arenosol (*Neossolo Quartzarênico*); SNap: Abruptic Solonetz (*Planossolo Nátrico*); MD: Mobile dunes; SB: Sand beach; UA: User's accuracy; PA: Producer's accuracy.

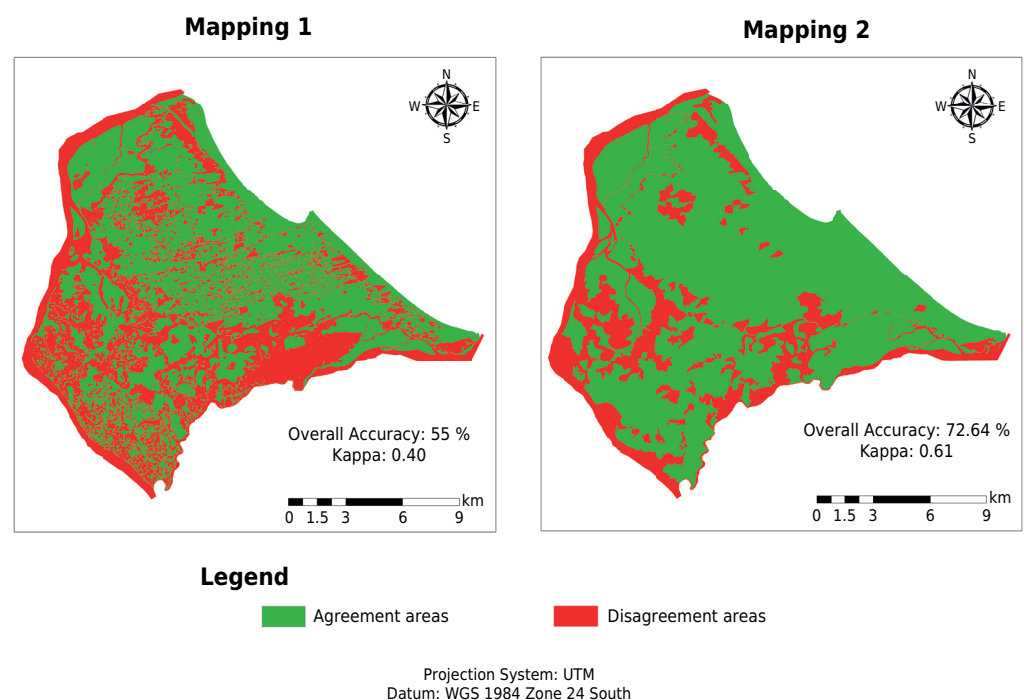


Figure 8. Analysis of agreement between the pre-existing soil map and the digital mappings.

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

GLEu: Eutric Gleysol (*Gleissolo Háplico*); GLti: Thionic Gleysol (*Gleissolo Tiomórfico*); AREu: Eutric Arenosol (*Neossolo Quartzarênico*); SNap: Abruptic Solonetz (*Planossolo Nátrico*); GL: Soils with glyec characteristics; AR: Soils with sand texture; FL: Soils with fluvic materials; MD: Mobile dunes; SB: Sand beach; UA: User's accuracy; PA: Producer's accuracy.

When validated by the traditional map, the map generated based on matrix 2 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.

CONCLUSIONS

The classification algorithm used, J48, showed satisfactory results at the mapping scale employed (1:100.000), demonstrating great potential to support soil cartography, since the algorithm selects and defines each attribute's priority during classification. Compared with the pre-existing map, the maps generated by Decision Trees showed a good 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 Digital Elevation Model - DEM/SRTM data proved to be low for the area, probably due to the low elevation of the study area, which may have contributed to the reduction of the accuracy of the whole set of data, according to the result of processing the matrices.




Digital maps of the Delta do Parnaíba River, despite having presented results ranging from reasonable to good, have limitations related to the distribution of data and the coverage of geographic space, since access to certain areas was limited. So, there is room for future improvements of the final products. Mapping units and soil groups were predicted at the highest possible quality and are sufficient for the local administration and academic research needs. For an environmental diagnosis at the regional scale, it would be necessary, among other efforts, to include other predictive variables and apply other pedometric methodologies.






ACKNOWLEDGEMENTS

The authors especially thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (309119/2020-0; 162224/2020-6; 443176/2014-0; G.S. Valladares), and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES for Scholarship (J. V. A. Amorim).





AUTHOR CONTRIBUTIONS






Conceptualization:  Gustavo Souza Valladares (equal),  João Victor Alves Amorim (lead) and  Mirya Grazielle Torres Portela (supporting).



Data curation:  Gustavo Souza Valladares (equal),  João Victor Alves Amorim (lead) and  Mirya Grazielle Torres Portela (supporting).



Formal analysis:  Andréa Maciel Lima (supporting),  Gustavo Souza Valladares (equal),  João Victor Alves Amorim (supporting),  Marcos Gervasio Pereira (lead) and  Mirya Grazielle Torres Portela (equal).



Funding acquisition:  Gustavo Souza Valladares (lead).

Investigation:  Andréa Maciel Lima (supporting),  Gustavo Souza Valladares (equal),  João Victor Alves Amorim (lead) and  Mirya Grazielle Torres Portela (supporting).





Methodology:  Andréa Maciel Lima (supporting),  Gustavo Souza Valladares (equal),  João Victor Alves Amorim (lead),  Marcos Gervasio Pereira (supporting) and  Mirya Grazielle Torres Portela (supporting).





Project administration:  Gustavo Souza Valladares (equal) and  João Victor Alves Amorim (equal).

Resources:  Gustavo Souza Valladares (equal) and  João Victor Alves Amorim (equal).






Software:  Gustavo Souza Valladares (supporting) and  João Victor Alves Amorim (lead).

Supervision:  Gustavo Souza Valladares (equal),  Marcos Gervasio Pereira (supporting) and  Mirya Grazielle Torres Portela (equal).

Validation:  Gustavo Souza Valladares (equal),  João Victor Alves Amorim (lead),  Marcos Gervasio Pereira (supporting) and  Mirya Grazielle Torres Portela (supporting).

Visualization:  Andréa Maciel Lima (supporting),  Gustavo Souza Valladares (equal),  Marcos Gervasio Pereira (supporting) and  Mirya Grazielle Torres Portela (equal).

Writing - original draft:  Andréa Maciel Lima (supporting),  João Victor Alves Amorim (lead) and  Mirya Grazielle Torres Portela (supporting).

Writing - review & editing:  Andréa Maciel Lima (supporting),  Gustavo Souza Valladares (lead),  João Victor Alves Amorim (equal),  Marcos Gervasio Pereira (supporting) and  Mirya Grazielle Torres Portela (supporting).

REFERENCES

- Alves Sobrinho T, Oliveira PTS, Rodrigues DBB, Ayres FM. Automatic delimitation of watershed using data SRTM. *Eng Agric.* 2010;30:46-57. <https://doi.org/10.1590/S0100-69162010000100005>
- Barros FML, Muehe D. Avaliação local da vulnerabilidade e riscos de inundação na zona costeira da Região dos Lagos, Rio de Janeiro. *Quat Environ Geosci.* 2010;2:55-66. <https://doi.org/10.5380/abequa.v2i1-2.14106>
- Barsi JA, Lee K, Kvaran G, Markham BL, Pedelty JA. The spectral response of the landsat-8 operational land imager. *Remote Sens.* 2014;6:10232-51. <https://doi.org/10.3390/rs61010232>
- Brasil. Exército Brasileiro. Diretoria de Serviço Geográfico. Banco de Dados Geográficos do Exército. Folha Topográfica Parnaíba, Folha SA 24 Y-A-IV, escala 1:100,000. Brasília, DF: BDGEx; 1972. Available from: <https://bdgex.eb.mil.br/mediador/>

- Brasil. Lei Nº. 9,985 de 18 de Julho de 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; Brasília, DF: Casa Civil, Subchefia para Assuntos Jurídicos; 2000.
- Brasil. Instituto Nacional de Meteorologia - Normais Climatológicas de Precipitação (1981/2010). Brasília, DF: Inmet; 2018. Available from: <https://portal.inmet.gov.br/normais>
- Brungard CW, Boettinger JL, Duniway MC, Wills AS, Edwards Jr TC. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*. 2015;239-240:68-83. <https://doi.org/10.1016/j.geoderma.2014.09.019>
- Cabral LJRS. Levantamento pedológico da Planície do Delta do Parnaíba, PI [dissertation]. Teresina: Universidade Federal do Piauí; 2018.
- Chagas CS. Digital mapping of soils by environmental correlation and neural networks in a watershed in the sea hills area [thesis]. Viçosa, MG: Universidade Federal de Viçosa; 2006.
- Chagas CS, Fernandes Filho EI, Vieira CAO, Schaefer CEGR, Carvalho Júnior W. Topographic attributes and Landsat7 data in the digital soil mapping using neural networks. *Pesq Agropec Bras*. 2010;45:497-507. <https://doi.org/10.1590/S0100-204X2010000500009>
- Costa JLPO, Cavalcanti APB. Phytogeography of the delta plain of the Parnaíba River, Piauí/Maranhão - Brasil: Analysis of the distribution of species and anthropogenic interference. *Observatorum*. 2010;2:84-92. <https://seer.ufu.br/index.php/Observatorum/article/view/44199>
- Dias LM, Coelho RM, Valladares GS, Assis ACC, Ferreira EP, Silva RC. Soil class prediction by data mining in an area of the sedimentary São Francisco basin. *Pesq Agropec Bras*. 2016;51:1396-404. <https://doi.org/10.1590/s0100-204x2016000900038>
- Gao BC. NDWI: A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens Environ*. 1996;58:257-66. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Guedes HAS, Silva DD. Comparison between hydrographically conditioned digital elevation models in the morphometric characterization of watersheds. *Eng Agric*. 2012;32:932-43. <https://doi.org/10.1590/S0100-69162012000500012>
- Hengl T, Rossiter DG. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. *Soil Sci Soc Am J*. 2003;67:1810-22. <https://doi.org/10.2136/sssaj2003.1810>
- Heung B, Ho HC, Zhang J, Knudby A, Bulmer CE, Schimdt MG. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*. 2016;265:62-77. <https://doi.org/10.1016/j.geoderma.2015.11.014>
- Höfig P, Giasson E, Vendrame PRS. Digital soil mapping based on map extrapolation between physiographically similar areas. *Pesq Agropec Bras*. 2014;49:958-66. <https://doi.org/10.1590/S0100-204X2014001200006>
- Horn BKP. Hill shading and the reflectance map. *Proc IEEE*. 1981;69:14-47. <https://doi.org/10.1109/PROC.1981.11918>
- Jensen JR. Sensoriamento remoto do ambiente: Uma perspectiva em recursos terrestres. São José dos Campos: Parêntese; 2009.
- Kempen B, Brus DJ, Stoorvogel JJ, Heuvelink GBM, Vries F. Efficiency comparison of conventional and digital soil mapping for updating soil maps. *Soil Sci Soc Am J*. 2012;76:2097-115. <https://doi.org/10.2136/sssaj2011.0424>
- Kheir RB, Greve MH, Abdallah C, Dalgaard T. Spatial soil zinc content distribution from terrain parameters: A GIS-based decision-tree model in Lebanon. *Environ Pollut*. 2010;158:520-8. <https://doi.org/10.1016/j.envpol.2009.08.009>
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-74. <https://doi.org/10.2307/2529310>
- Marandola Júnior E, Marques C, Paula LT, Cassaneli LB. Crescimento urbano e áreas de risco no litoral norte de São Paulo. *Rev Bras Estud Popul*. 2013;30:35-56. <https://doi.org/10.1590/S0102-30982013000100003>

- Marengo JA, Torres RR, Alves LM. Drought in Northeast Brazil- past, present, and future. *Theor Appl Climatol*. 2017;129:1189-200. <https://doi.org/10.1007/s00704-016-1840-8>
- McBratney AB, Odeh IOA, Bishop TFA, Dunbar MS, Shatar TM. An overview of pedometric techniques for use in soil survey. *Geoderma*. 2000;97:293-327. [https://doi.org/10.1016/S0016-7061\(00\)00043-4](https://doi.org/10.1016/S0016-7061(00)00043-4)
- Mendonça F, Danni-Oliveira IM. *Climatologia: Noções básicas e climas do Brasil*. São Paulo: Oficina de Textos; 2007.
- Minasny B, McBratney AB. Digital soil mapping: A brief history and some lessons. *Geoderma*. 2015;264:301-11. <https://doi.org/10.1016/j.geoderma.2015.07.017>
- Moore ID, Gessler PE, Nielsen GA, Peterson GA. Soil attribute prediction using terrain analysis. *Soil Sci Soc Am J*. 1993;57:443-52. <https://doi.org/10.2136/sssaj1993.03615995005700020026x>
- Nolasco-Carvalho CC, Franca-Rocha W, Uchoa JM. Digital soil map: a methodological proposal using fuzzy inference. *Rev Bras Eng Agr Amb*. 2009;13:46-55. <https://doi.org/10.1590/S1415-43662009000100007>
- Portela MGT, Espindola GM, Valladares GS, Amorim JVA, Frota JCO. Vegetation biomass and carbon stocks in the Parnaíba River Delta, NE Brazil. *Wetlands Ecol Manage*. 2000;28:607-22. <https://doi.org/10.1007/s11273-020-09735-y>
- Rennó CD, Nobre AD, Cuartas LA, Soares JV, Hodnett MG, Tomasella J, Waterloo MJ. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sens Environ*. 2008;112:3469-81. <https://doi.org/10.1016/j.rse.2008.03.018>
- Rizzo R, Demattê JAM, Lepsch IF, Gallo BC, Fongaro CT. Digital soil mapping at local scale using a multi-depth Vis-NIR spectral library and terrain attributes. *Geoderma*. 2016;274:18-27. <https://doi.org/10.1016/j.geoderma.2016.03.019>
- Rosa AS. *Funções de predição espacial de propriedades do solo [dissertation]*. Santa Maria: Universidade Federal de Santa Maria; 2012.
- Ruiz LFC, ten Caten A, Dalmolin RSD. *ADtoSIG Conversion Software*. Santa Maria: Universidade Federal de Santa Maria; 2011.
- Santos-Filho FS, Almeida Junior EB, Soares CJRS, Zickel CS. Faces of the Salt Marshes of Delta Parnaíba, Northeastern Brasil. *Rev Bras Geograf Fis*. 2010;3:218-27. <https://doi.org/10.26848/rbgf.v3i3.232605>
- Silveira CT, Oka-Fiori C, Santos LJC, Sirtoli AE, Silva CR, Botelho MF. Soil prediction using artificial neural networks and topographic attributes. *Geoderma*. 2013;195-196:165-72. <https://doi.org/10.1016/j.geoderma.2012.11.016>
- Sirtoli AE, Silveira CT, Montovani LE, Sirtoli ARA, Oka-Fiori C. Landscape attributes derived from digital elevation model and relationships with soils. *Sci Agrar*. 2008;9:317-29. <https://doi.org/10.5380/rsa.v9i3.11517>
- Sousa RS. *Planície Costeira do Estado do Piauí: Mapeamento das unidades de paisagem, uso e cobertura da terra e vulnerabilidade ambiental [dissertation]*. Teresina: Universidade Federal do Piauí; 2015.
- Taghizadeh-Mehrjardi R, Nabiollahi K, Kerry R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma*. 2016;266:98-110. <https://doi.org/10.1016/j.geoderma.2015.12.003>
- ten Caten A. *Digital soil mapping: Methods to meet the demand for soil spatial information [thesis]*. Santa Maria: Universidade Federal de Santa Maria; 2011.
- Valeriano MM. Dados topográficos. In: Florenzano TG. *Geomorfologia, conceitos e tecnologias atuais*. São Paulo: Oficina de Textos; 2008. p. 72-104.
- Valladares GS, Cabral LJRS. Mapeamento geológico da Planície do Delta do Parnaíba-PI. In: 27th Symposium of Northeastern Geology; 2017 Nov 12-15; João Pessoa, Paraíba. João Pessoa: Universidade Federal de João Pessoa; 2017. Available from: http://www.geologiadordeste.com.br/anais/index_sessao.php#resumos

Witten IH, Frank E. Data mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann; 2005.

Witten IH, Frank E, Hall MA. Data mining: Practical machine learning tools and techniques. 3rd ed. San Francisco: Morgan Kaufmann; 2011.

Žížala D, Minařík R, Skála J, Beitlerová H, Juřicová A, Rojas JR, Penížek V, Zádorová T. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. *Catena*. 2022;212:106024. <https://doi.org/10.1016/j.catena.2022.106024>