Volume 22 Issue 38
Originally issued as B3



Global Journal of human-social science: b
Geography, Geo-Sciences, Environmental Science & Disaster Management
Volume 22 Issue 3 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 2249-460x & Print ISSN: 0975-587X

Clustering of Fine-Grained Tropical Soils using Data Science Tools Applied to their Geotechnical Properties

By Mayssa Alves Da Silva Sousa, Roberto Quental Coutinho & Laura Maria Goretti Da Motta
State University of Maranhão

Abstract- The characterization of fine-grained tropical soils for use in pavements has evolved since the 1980s, however, even today these soils are still discarded or underused in infrastructure works because they do not fully meet the requirements established by traditional classification methodologies or even by the CBR. Tropical soils present peculiarities of geotechnical behavior regarding elastic and plastic deformability, as many authors have already observed. This article contributes to this distinction by analyzing the grouping of thirteen fine-grained soils from northeastern Brazil through the application of data science tools to the results of geotechnical tests. More than fifty geotechnical parameters obtained in the laboratory were considered. By means of simple and multiple linear regressions, they were analyzed in a hierarchical cluster, using Ward's linkage method and Euclidean distance. The results showed that the mechanical behavior of soil compaction and the granulometry, especially the quantities of silt and fine sand, were decisive for the initial division of soils into clusters.

Keywords: *tropical soils. geotechnical properties. geotechnical behavior. cluster analysis. data science.*

GJHSS-B Classification: *DDC Code: 418.007 LCC Code: P53*



Strictly as per the compliance and regulations of:



© 2022. Mayssa Alves Da Silva Sousa, Roberto Quental Coutinho & Laura Maria Goretti Da Motta. This research/review article is distributed under the terms of the Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). You must give appropriate credit to authors and reference this article if parts of the article are reproduced in any manner. Applicable licensing terms are at https://creativecommons.org/licenses/by-nc-nd/4.0/.

Clustering of Fine-Grained Tropical Soils using Data Science Tools Applied to their Geotechnical Properties

Mayssa Alves Da Silva Sousa [α], Roberto Quental Coutinho [σ] & Laura Maria Goretti Da Motta [ρ]

Abstract- The characterization of fine-grained tropical soils for use in pavements has evolved since the 1980s, however, even today these soils are still discarded or underused in infrastructure works because they do not fully meet the requirements established by traditional classification methodologies or even by the CBR. Tropical soils present peculiarities of geotechnical behavior regarding elastic and plastic deformability, as many authors have already observed. This article contributes to this distinction by analyzing the grouping of thirteen fine-grained soils from northeastern Brazil through the application of data science tools to the results of geotechnical tests. More than fifty geotechnical parameters obtained in the laboratory were considered. By means of simple and multiple linear regressions, they were analyzed in a hierarchical cluster, using Ward's linkage method and Euclidean distance. The results showed that the mechanical behavior of soil compaction and the granulometry, especially the quantities of silt and fine sand, were decisive for the initial division of soils into clusters. For refining the clusters, we considered the different types of geotechnical parameters (physical, chemical and mechanical) in order to highlight other important variables in the distinction of the behavior of these materials. Additionally, microscopic images were compared also using data science techniques and the results corroborated those obtained in cluster analysis, indicating that soil mineralogical characteristics are associated with with the results of laboratory geotechnical tests.

Keywords: tropical soils. geotechnical properties. geotechnical behavior. cluster analysis. data science.

I. Introduction

In Brazil, the study of the behavior of fine-grained tropical soils, in the perspective of pavement engineering, began several decades ago, especially with the works of Nogami and Villibor, 1991, 1995), which introduced the classification called MCT. These are soils typical of the tropical environment that, according to the concept adopted by the *Committee on Tropical Soils of the International Society of Soil Mechanics and Foundation Engineering - ISSMFE* in 1985, present "peculiarities of properties and behavior, in relation to non-tropical soils, due to the performance in the same geological and/or pedological processes, typical of humid tropical regions".

It is also emphasized that the introduction of repeated load tests for the determination of the resilient modulus and permanent deformation of this type of soil (Medina and Preussler, 1980; Svenson, 1980) consolidated the appropriate mechanical characteristics of fine-grained tropical soils, complementing the MCT methodology.

Due to fine granulation, most of these soils are usually discarded or underused in infrastructure works because they do not present geotechnical parameters that fit the traditional selection criteria (*Transportation Research Board - TRB and California Bearing Ratio - CBR*). However, it has been demonstrated by several surveys (Nogami and Villibor, 1991; Guimarães, 2009; Medina and Motta, 2015; Sousa, 2016; Dalla Roza and Motta, 2018; Lima, Motta and Guimarães, 2017; Lima et al., 2020; Guimarães, Motta and Castro, 2019; Guimarães, Silva Filho and Castro, 2021; among several others) that, regardless of granulometry, consistency indexes and CBR, many of the fine-grained tropical soils have excellent mechanical performance in terms of resilience and plastic deformability, justifying their use in road and railway pavements.

Currently, many laboratory tests are carried out to expand knowledge about soil behavior for geotechnical purposes, determining their physical, mechanical, chemical and mineralogical characteristics. The joint analysis of the results obtained by these tests can be performed by means of clustering techniques and can provide valuable information to understand the behavior of soils considering several variables at the same time, by a multivariate analysis.

In this context, Frank and Todeschini (1994) define Cluster Analysis as a set of multivariate exploratory methods that seek to find clusters, based on some criterion of similarity between objects (or variables), and that the result of clustering depends greatly on the method used, the standardization of variables and the measure of similarity chosen. The main premise is that the groups or "clusters" formed should be as homogeneous as possible and the differences between the various clusters as large as possible.

Author α: Civil Engineering Department, State University of Maranhão, Brazil. e-mail: mayssasousa@professor.uema.br
Author σ: Civil Engineering Department, Federal University of Pernambuco, Recife, Brazil. e-mail: roberto.coutinho@ufpe.br
Author ρ: Civil Engineering Department, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. e-mail: laura@coc.ufrj.br

© 2022 Global Journals

In data science, cluster analysis is part of unsupervised Machine Learning technology that encompasses a set of tools for creating clusters with homogeneous properties from a large number of heterogeneous samples.

Data science according to VanderPlas (2016) is difficult to define, but can be considered as an interdisciplinary set of skills that are becoming increasingly important in many applications in industry and academia and comprises the intercession of three distinct overlapping areas: 1 – *Math & Statistics Knowledge*, to model and summarize data sets; 2 – *Hacking Skills*, to design and use algorithms to store, process and visualize this data efficiently and, 3 – *Substantive Expertise*, knowledge necessary to interpret the results.

For the application of Data Science it is common to use programming languages such as R and Python (among others) through the implementation of codes and sets of functions (libraries) that allow to manipulate and treat data, as well as generate relevant information about the data in seconds.

It is noteworthy that the Python programming language has established itself as one of the most popular languages for scientific computing due to its interactive nature and its system of maturation of scientific libraries, being an attractive choice for the development of algorithms and exploratory data analysis (Millman and Aivazis, 2011).

There are several libraries with several modules with different functionalities and are used at different stages of the analysis in Data Science with Python, whose focus varies according to the analyst's objective. Thus, to deepen the knowledge about the main libraries and modules, the following references are recommended (not limited to these): Millman and Aivazis (2011); Pedregosa (2011); Harris et al., (2020) and Virtanen et al., (2020) and the book "Data Science of Zero" by Joel Grus (2016) that present concepts and details on the subject.

There are several clustering methods as shown Hastie, Tibshirani and Friedman (2009), Hair et al., (2009), Härdle and Simar (2015), Forsyth (2018), among others. The hierarchical method is the one used in this article, because it is the most frequently applied in practice. Härdle and Simar (2015), indicates that it starts with the best possible structure, calculates the distance matrix for the clusters, and joins the clusters that have the shortest distance.

However, it should be emphasized that clustering techniques in general, and especially hierarchical clustering, is an exploratory analysis of data, and different combinations may reveal different characteristics of the data set, as analyzed by Chen et al., (2007).

Härdle and Simar (2015) say that cluster analysis can be divided into two fundamental steps: 1 – Choice of proximity measure (each pair of observations is verified as to the similarity of their values) and, 2 – Choice of the cluster creation algorithm (based on proximity measurements, objects are assigned to clusters so that the differences between them become large and observations within the cluster become as approximate as possible).

The proximity between the data is measured by a distance or matrix of similarity distances whose components provide the coefficient of similarity or the distance between two points. There is a variety of distance measurements for the various types of data, and for quantitative variables, the most used are the Euclidean (used in this article), generalized/weighted, and Minkowski distance.

There is also a variety of cluster linkage methods, the main ones are indicated in the Table 1. Hair et al., (2009) state that, combined with the chosen measure of similarity, the clustering algorithm provides the means to represent the similarity between clusters with multiple members. However, according to Härdle and Simar (2015), Metz (2006) and Frank and Todeschini (1994) there is no "correct" combination of distance measurement and linkage method.

Table 1: Linkage methods in cluster analysis

Linkage methods	Clustering shapes	Comment
Single linkage	Defines the distance between two groups as the shortest distance between an element of one group to an element of the other group, also called the Nearest Neighbour algorithm.	Tends to produce large clumps, weakly linked and with little internal cohesion.
Complete linkage	The distance between two clusters is calculated as the greatest distance between two objects in opposite clusters, also called the Farthest Neighbour algorithm.	Tends to produce well separated and small clumps.
Average linkage	The distance between two clusters is calculated as the average of the distances between all pairs of objects in opposite clusters.	It proposes a compromise between the two previous algorithms.

© 2022 Global Journals

Centroid linkage	The distance between two clusters is calculated as the distance between their centroids.	Each cluster is represented by its centroid.
Ward method	The distance between two clusters is calculated by summing the squared deviations of each object from the centroid of its own cluster, joining two clusters that result in the smallest increase in the sum of squares of the total error within the group, also called the least variance method.	Unlike the other methods, brings together groups that do not dramatically increase heterogeneity, in this way, it unifies the groups so that the variation within these groups is minimized, groups created as homogeneous as possible.

Note: developed on the basis of Frank and Todeschini (1994) and Härdle and Simar (2015).

Another important detail for cluster analysis includes the standardization of the variables, since most cluster analyses using distance measurements are very sensitive to different scales or magnitudes between variables. In general, variables with higher dispersion (higher standard deviations) have a greater impact on the final similarity value. The most common form of standardization is the conversion of each variable into standard scores (Z scores) by subtraction of the mean and division by the standard deviation for each variable. The process converts each initial data score into a standardized value with an average of 0 and a standard deviation of 1, eliminating the bias that is introduced by the differences in the scales of the various attributes or variables used in the analysis (Hair et al., 2009).

The result of Cluster Analysis can be presented in the form of a graph, called dendrogram, where the observations, the sequence of the clusters and the distances between the clusters are presented. Hastie, Tibshirani and Friedman (2009) state that a dendrogram provides a complete interpretative description of the hierarchical cluster in a graphical format and that this is one of the main reasons for the popularity of this clustering method.

The main objective of this article is to analyze the clustering of fine-grained tropical soils in relation to their geotechnical properties in association with physical, mechanical, chemical and mineralogical characteristics, using data science tools (the Python programming language). In addition to the analysis and discussion of the hierarchical cluster dendrogram, the article compares microscopic images of soils in order to identify the similarity between them in mineralogical terms, verifying the similarity of the comparison with the result of the cluster analysis.

II. Materials and Methods

a) *Tropical Soils Studied*

Thirteen fine-grained soils were studied (with maximum of 10% of retained material in the no. 10 sieve - with a 2.0 mm opening, according to the criteria of the MCT Methodology: M - Miniature, C - Compacted, T - Tropical), collected in horizon B of road or deposit areas available in the Metropolitan Region of Recife, which is composed of 15 municipalities (including the capital – Recife), as indicated in Figure 1.

The MCT Methodology, created by Job Shuji Nogami and Douglas Fadul Villibor in 1980, allows the initial classification of soils into two large groups (lateritic behavior - L and Non-lateritic - N) and the categorization of these into the classes: LA - Lateritic Sand; LA' - Lateritic Sandy Soil; LG' – Lateritic Clay soil; NA - Non-Lateritic Sand; NA' - Non-Lateritic Sandy Soil; NS' – Non-Lateritic Silty Soil and NG' – Non-Lateritic Clay soil, more detailed in Villibor and Nogami (2009).

© 2022 Global Journals

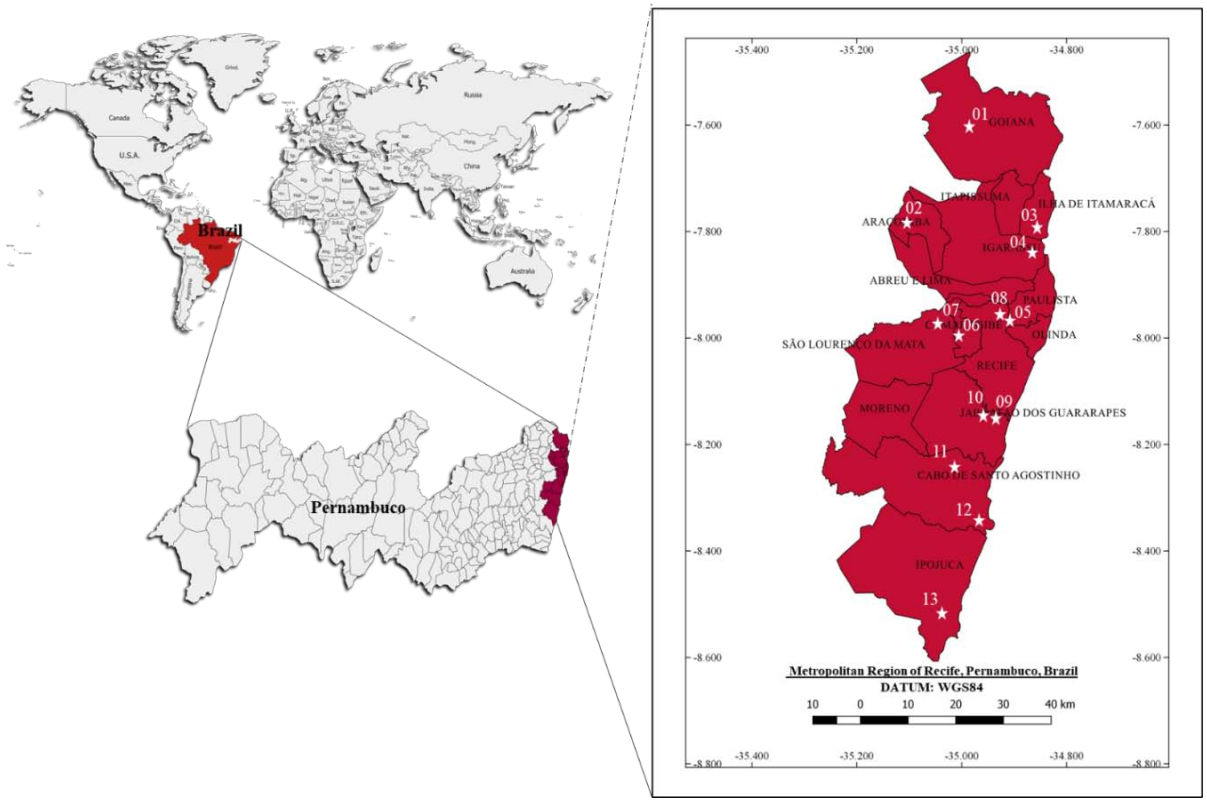


Figure 1: Location of the soils in the Metropolitan Region of Recife - Pernambuco, Brazil

The characteristic climate of the region is marked by large volumes of rainfall (1,500 to 2,000 mm per year) and high temperatures (> 20°C), according to the Agência Pernambucana de Água e Clima (APAC), a typical condition of tropical environments according to CHESWORTH et al., (2008).

Most of the selected materials are from the Barreiras Geological Formation which is formed by unconsolidated sediments with diversified particle size, observing the predominance of clay and sandy clay sediments and, less frequently, sandy, and may also present ferruginous concretions and pebbles (CPRM, 2014 and Coutinho et al., 2019).

In terms of pedology, the predominant soils are yellow acrisols and yellow latosols (in the Brazilian classification), which in the international soil classification system of the Food and Agriculture Organization of the United Nations (FAO) can comprehend the classes of ferralsols and acrisols.

They are soils that have suffered a lot from weathering processes and often present the phenomenon of natural cohesion/cementation, which is a strong pedogenetic hardening of the soil when it reaches the dry state (reversible in the wet state) due to the high concentrations of Fe hydroxides and Al oxides (CPRM, 2014; Guimarães, Silva Filho and Castro, 2021; Sousa et al., 2021; Coutinho and Sousa, 2021), as was clearly identified in the soils studied in this article (examples in Figure 2).



(a) 12SU-LG'

(b) 13IP-LA'

Figure 2: Examples of the natural cementation of the thin tropical soils of this research

© 2022 Global Journals

b) Geotechnical Tests And Parameters

The geotechnical parameters of the soils were obtained by physical, mechanical, chemical and mineralogical tests in the laboratory following, whenever possible, the Brazilian standards of the National Department of Transport Infrastructure (DNIT, formerly DNER). In all, 57 variables were considered, and Table 2 presents a list of the tests performed with an indication of which of the parameters obtained through them were used as the basis for the generation of soil clustering, including their respective units, as well as the standards and manuals that served as a reference for the execution of each test.

Table 2: Performed tests, obtained geotechnical parameters and reference standards/manuals

Test	Parameters	Standards
Real Particle Density	Real Particle Density (δ, g/cm³)	DNER-ME 093/94
Atterberg limits	Liquid Limit (LL, %), Plastic Limit (PL, %), Plasticity Index (PI, %)	DNER-ME 122/94; DNER-ME 082/94
Granulometric analysis (with deflocculant)	%Clay, %Silt, %Sand, %Fine Sand (%FS), %Medium Sand (%MS), % Coarse Sand (%CS), %Gravel(%Grav)	DNER-ME 051/94
MCT Classification	Coefficients c' and d', e' index, Immersion Mass Loss (PMI, %), Immersion Mass Loss – specimen with optimal soil moisture (PMIwo, %)	DNER-ME 258/94; DNER-ME 256/94
Compaction	Apparent dry specific weight ($\rho_{dmáx}$, g/cm³) and optimal moisture content (w_o, %)	DNIT 164/2013-ME
California Bearing Ratio	CBR (%) and expansion (Exp., %) at the optimal moisture content	DNIT 172/2016-ME
Chemical tests	hydrogen potential in: water (pH_{H2O}) and KCl (pH_{KCl}), ΔpH (pH_{KCl} - do pH_{H2O}), Organic Matter (M.O., $g.kg^{-1}$), Base saturation (V, %), Aluminum saturation (S, %), Cation Exchange Capacity (CEC, $cmol_c dm^{-3}$)	EMBRAPA (2017)
Soil-Water Characteristic Curve	Soil suction at optimal moisture contente (S_{wo}, kPa), saturation humidity (θs, %), residual moisture (θ_r, %), suction obtained in the resilient module specimens (SMR, kPa) and parameters of Model's Gitirana Jr. and Fredlund (2004) - Ψb1, Ψres1, Sres1, Ψb2, Sb, Ψres2, Sres2	ASTM D5298 – 16
Resilient Modulus	Resilient Modulus medium (MR_{medium}, MPa), regression coefficients of linear models as a function of stresses: confining stress (K1TC e K2TC) and deviator (K1TD e K2TD), Composite model parameters (K1MC, K2MC e K3MC)	DNIT 134/2018-ME
Permanent Deformation	Maximum permanente deformation (ε_p, mm) obtained under confining stress - σ_3 (120 kPa) and deviator stress - σ_d (360 kPa), Parameters of the Guimaraes Model (2009) - Ψ1, Ψ2, Ψ3 and Ψ4	DNIT 179/2018-IE

The LL and PL tests were performed both with samples previously dried in the air and destroyed (standard procedure) and without previous drying, and the parameters obtained were identified as follows: without previous drying - LL1, PL1 and PI1 and, with drying and destruction: LL2, PL2 and PI2.

The test of the soil-water characteristic curve was performed using the filter paper method, with the specimen compacted on the intermediate Proctor energy and considering a drying and wetting mixed trajectory. The model adjustment parameters of Gitirana Jr. and Fredlund (2004) follow equation 1, applicable to soils with bimodal behavior commonly identified in lateritic tropical soils. The coefficients of determination (R^2) obtained were all very close to 1 (>0.91) showing suitability of the model to the soil behavior of this article.

$$S_e = \frac{S_1 - S_2}{1+\left(\frac{\psi}{\sqrt{\psi_{b1}\psi_{res1}}}\right)^{d1}} + \frac{S_2 - S_3}{1+\left(\frac{\psi}{\sqrt{\psi_{res1}\psi_{b2}}}\right)^{d2}} + \frac{S_3 - S_4}{1+\left(\frac{\psi}{\sqrt{\psi_{b2}\psi_{res2}}}\right)^{d3}} \quad (1)$$

Where: S = degree of saturation; = suction obtained in the laboratory test;b = suction at the air inlet point; res = residual suction and d1, d2, d3 = model's parameters.

Mechanical compaction, CBR, resilient modulus (MR) and permanent deformation tests were performed on the intermediate Proctor energy. The equations that were used to model resilient behaviors and permanent soil deformation are indicated in Table 3.

© 2022 Global Journals

Table 3: Models used to express the resilient behavior and permanent deformation of the soils of this research

References	Equation	Variables	Parameters
Hicks and Monismith (1970)	MR=k13k2	3	k1 and k2
Svenson (1980)	MR=k1dk2	d	k1 and k2
Macêdo (1996)	MR=k13k2dk3	3;d	k1, k2 and k3
Guimarães (2009)	p(%)=1302d03N4	3;d and N	1, Ψ2, 3 and 4

Notation: 3 = confining stress; d= deviator stress; p(%) = Specific permanent deformation, N = number of loading cycles and, k1, k2, k3, 1, Ψ2, 3 and 4 are regression parameters, 0 = reference stress, considered with the atmospheric pressure of 100 kPa.

c) *Data Science Tools For Cluster Analysis*

The implementation of the codes required for the analyses was carried out in the Python programming language (version 3) available in the Anaconda virtual environment (https://www.anaconda.com/). In addition to the standard Python library, modules and functions from other libraries were used as indicated in Table 4, which explains the general objectives of each (in the application column) and in the column "functions and modules" is the specific indication of the main tools that were used.

Table 4: Main Libraries, modules and functions used in the analysis with Python in this research

Libraies	Aplication	Modules and functions
Pandas https://pandas.pydata.org/	Data manipulation and analysis: structures and operations to manipulate numerical tables and time series.	read_excel, drop, set_index, head, shape
Numpy https://numpy.org/	Data analysis: mathematical functions, random number generators, linear algebra routines, Fourier transformations, etc.	Array, arrange, mean, std, argsort
Scipy https://www.scipy.org/	Data modeling: fundamental algorithms for statistical functions (probability distributions, hypothesis testing, frequency statistics, correlation functions, etc.).	scipy.cluster, hierarchy, hierarchy.linkage (method='ward',metric='euclidean'), hierarchy.dendrogram
Scikit-learn https://scikit-learn.org/	Data Modeling: Machine learning algorithms, supervised and unsupervised (Classification, Regression, Clustering, Model Selection, etc.)	sklearn.preprocessing, StandardScaler, transform
Scikit-image https://scikit-image.org/	Image processing: functions for manipulating scientific, specific or general-purpose images, operations on Numpy matrices, manipulation of exposure and color channels, detection and segmentation of objects.	Feature
Imageio https://pypi.org/project/imageio/	Image manipulation: reading and writing image data, including animated images, volumetric data, and scientific formats.	imread
Matplotlib https://matplotlib.org/	Data presentation and exploration: creating graphs and general data visualization	plot.figure, plot.title, plot.xlabel, , plot.yabel, subplot , imshow

For the analysis of data clusters through the hierarchical method and implementation via codes and libraries in Python, the data were standardized by applying the Z-score Normalization Method, and for this an array was created through the "array" function of the Numpy library and the functions "StandardScaler" and "transform" of the Scikit-learn library were applied.

Then, to obtain hierarchical cluster of the (standardized) datasets, the hierarchy function of the Cluster module of the SciPy library was imported. The implementation was made considering Ward's linkage method and Euclidean distance.

© 2022 Global Journals

d) Image Comparison

As a complementary analysis and a way of using data science tools for image analysis, images were separated from the fine-earth fraction of soils (in this case the material passing in sieve no. 10 and retained in no. 200) made in the Stereo Microscope Zeiss Discovery V8 in order to compare the soils to each other and to verify some similarity with the results obtained in the clustering.

The mineralogy of the fraction passing in sieve no. 200 was studied by performing the X-ray Diffraction test and images of the coarse fraction of the soil (retained in sieve no. 10) were also extracted, but were not considered in the analyses of this article.

In summary, the process of comparing images in Python is done by implementing a function that receives the image and processes it by transforming it into vectors. The function applied in this article was the BIC (Border-Interior Pixel Classification), which transforms and rescales the image, and finally computes two color histograms: one for the "interior" pixels and the other for the "border" pixels (edges of the image), then normalizes the histograms and concatenates them into a single vector. Vectors represent the frequencies of colors in the image, which allows, with the aid of a distance measurement, to compare images for color similarity.

To make the comparison, you must set an image as a reference so that the distance of the other images related to it is calculated. In this part of the analysis, we chose to use the same distance measurement used in the cluster analysis that was the Euclidean distance.

III. Results and Discussion

a) Geotechnical Parameters

Table 5 presents part of the database of the studied soils: geotechnical classifications (TRB, USCS and MCT), MCT classification, real density, Atterberg limits, granulometry, mechanical compaction parameters, CBR (and expansion), MR, DP and chemical data. Table 6 presents the continuity of the database with parameters associated with the characteristic curve of the soils. The units of measurement of each parameter were presented in Table 2.

Regarding the USCS class, five samples were classified as ML (silt soils of low compressibility), four were classified as SC (clay sand), and two as SM (silty sand). Soil 03 was classified as SM-SC and soil 13 was classified as CL (clay soil of low compressibility).

Table 5: Classification database, true density, consistency limits, granulometry, compaction, CBR, expansion, MR, DP and chemical data from soils studied in this research

Soil	1	2	3	4	5	6	7	8	9	10	11	12	13
MCT	LG'	LG'	NA'	LG'	LA'	LA'	NA'	LA'	LA'	LA'	LA'	LG'	LA'
TRB	A-6	A-7-6	A-4	A-6	A-6	A-6	A-6	A-4	A-4	A-7-6	A-7-5	A-6	A-6
USCS	ML	ML	SM-SC	SC	SC	SM	ML	ML	SM	SC	ML	SC	CL
d'	247	139	133	87	131	140	147	107	109	217	188	150	154
e'	0,94	0,62	1,19	0,93	0,62	0,62	1,16	0,84	0,91	0,87	0,84	0,81	0,81
c'	2,0	1,9	1,2	1,7	1,4	1,05	1,05	1,29	0,92	1,05	1,0	1,67	1,1
PMI	75	10	153	58	9	10	143	40	58	56	49	40	40
PMI_{wo}	75	0	98	38	9	0	34	0	0	0	0	39	42
δ	2,63	2,67	2,65	2,63	2,64	2,67	2,69	2,70	2,68	2,68	2,68	2,67	2,68
LL1	32,2	33,7	22,0	26,2	30,0	28,5	32,2	34,5	27,4	27,5	33,9	27,8	28,7
PL1	22,8	25,9	18,0	20,3	20,4	20,4	24,4	26,5	19,3	19,8	30,4	18,7	19,8
PI1	9,4	7,8	4,0	5,9	9,6	8,1	7,8	8,0	8,1	7,7	3,5	9,1	8,9
LL2	35,6	41,3	24,1	33,4	37,9	39,9	39,9	39,5	34,7	40,5	46,6	34,8	35,9
PL2	25,0	27,9	17,4	21,7	23,1	26,5	28,5	32,4	24,9	25,4	30,1	21,4	24,2
PI2	10,6	13,4	6,7	11,7	14,8	13,4	11,4	7,1	9,8	15,1	16,5	13,4	11,7
%Clay	50,2	47,8	32,3	38,8	41,7	42,6	46,5	47,0	42,0	40,3	45,5	34,9	44,7
%Silt	14,2	19,0	5,6	6,6	4,8	5,2	13,6	19,3	5,7	8,1	6,8	4,4	8,4
%FS	25,3	18,7	55,9	36,2	31,5	31,1	27,3	21,8	35,3	29,9	25,9	33,6	25,5
%MS	10,0	11,2	6,1	17,3	18,2	20,6	12,3	11,5	16,7	20,2	20,9	25,7	20,5
%CS	0,3	1,2	0,1	0,8	2,1	0,6	0,3	0,3	0,3	0,8	0,9	0,9	0,9
%Grav	0,1	2,1	0,0	0,2	1,7	0,1	0,0	0,1	0,0	0,8	0,1	0,4	0,1
$\rho_{dmáx}$	1,79	1,75	1,97	1,91	1,94	1,88	1,78	1,69	1,90	1,87	1,86	1,94	1,86
W_o	17,4	18,4	10,6	13,1	13,7	13,8	16,5	19,0	12,7	12,7	15,6	11,2	13,2

© 2022 Global Journals

CBR	12,8	15,0	35,0	15,4	25,0	17,0	21,0	15,2	30,0	16,0	27,5	29,0	23,0
Exp.	0,05	0,00	0,15	0,60	0,00	0,00	0,05	0,10	0,00	0,20	0,00	0,20	0,00
MR_{medium}	416	394	865	477	425	667	557	394	507	611	487	478	689
K1TC	222	144	200	175	153	164	231	214	519	345	222	359	227
K2TC	-0,24	-0,33	-0,48	-0,32	-0,33	-0,49	-0,28	-0,2	0,01	-0,19	-0,25	-0,09	-0,34
K1TD	205	144	246	171	146	177	227	197	405	327	202	297	243
K2TD	-0,33	-0,41	-0,52	-0,42	-0,44	-0,60	-0,36	-0,29	-0,09	-0,26	-0,36	-0,20	-0,41
K1MC	354	282	289	295	263	168	384	342	650	462	403	483	341
K2MC	0,33	0,32	0,03	0,25	0,35	0,02	0,29	0,30	0,34	0,21	0,39	0,35	0,14
K3MC	-0,51	-0,53	-0,50	-0,52	-0,62	-0,64	-0,51	-0,43	-0,32	-0,39	-0,55	-0,43	-0,47
ε_p	6,52	6,59	2,93	7,19	1,63	6,66	4,18	3,63	0,78	3,37	3,15	1,30	6,40
$\Psi 1$	0,21	0,28	0,13	0,46	0,07	0,02	0,03	0,01	0,08	0,10	0,04	0,09	0,47
$\Psi 2$	-0,45	0,79	0,25	-0,51	0,07	-2,31	-1,86	-1,22	0,04	0,17	-0,53	-0,27	1,09
$\Psi 3$	1,73	1,30	1,24	1,30	0,07	3,90	2,98	3,96	0,91	1,42	2,39	1,07	0,95
$\Psi 4$	0,06	0,06	0,06	0,05	0,07	0,05	0,06	0,08	0,05	0,07	0,06	0,05	0,05
pH_{H2O}	4,7	4,9	4,9	4,8	4,6	4,6	4,4	4,6	4,6	4,5	4,4	4,8	4,7
pH_{KCl}	4,0	4,1	4,0	3,9	3,9	4,0	4,1	4,1	4,0	3,9	4,0	4,0	4,0
ΔpH	-0,7	-0,8	-0,9	-0,9	-0,7	-0,6	-0,3	-0,5	-0,6	-0,6	-0,4	-0,8	-0,7
M.O.	7,59	7,12	3,85	5,53	5,11	4,37	4,32	7,19	5,68	6,25	6,67	5,94	4,43
CEC	4,63	4,55	2,80	3,55	3,75	3,85	3,84	4,40	5,42	5,74	4,46	4,07	4,64
S	36,6	28,4	43,4	61,5	79,1	54,3	41,4	53,2	73,1	69,0	56,6	86,3	55,5
V	25,1	30,6	28,9	19,4	8,8	17,9	17,7	13,4	7,2	9,4	10,3	3,9	15,7

In the TRB classification, three samples were classified as A-4 (non-plastic silt soils to moderately plastic), seven as A-6 (plastic clay soils), one (soil 11) as A-7-5 (PI moderate in relation to LL, and may be elastic and subject to high volume variation) and two (02 and 10) as A-7-6 (high PI in relation to LL and are subject to extremely high volume changes).

According to the TRB classification, all soils of the research would present poor to bad behavior as a subbed layer of pavements, however, considering the results of the Resilient Modulus (MR) and Permanent Deformation (PD) trials, the studied soils present excellent behaviors regarding these aspects, evidencing that fine-grained tropical soils present peculiar behavior, as noted in several works already mentioned.

Regarding the MCT classification, it is verified that in general, the samples further north of the metropolitan region of Recife present clayey behavior, except for sample 03 (fine sandy soil), moving to a more sandy behavior when collected in the south region (except for sample 12).

It is observed that the coefficient d', which is associated with the inclination of the dry branch of the Mini-MCV compaction curve (in MCT), presents important variation between the samples, indicating different behaviors. Villibor and Nogami (2009) specify that d' values above 20 indicate soils of lateritic behavior and above 100 (very high) refers to the typical behavior of fine clayey sands. It is noted that most of the samples evaluated present these characteristics in this aspect and also granulometric and mineralogical.

Regarding Mass Loss by Immersion (PMI_{wo}), the values obtained for specimens (CP) molded in optimum moisture were 0% to almost 50% of the soils, even after 24 hours of immersion. The e' index, which associates the MLI and the coefficient d', reflects the lateritic or non-lateritic behavior of the analyzed soil. Soils with low mass loss and high d' values result in lower values of e' and more evident lateritic behavior: 02, 05, 06, 08 to 13.

As for the values of c', which is associated with the slope of the soil deformability curve, it is understood that the higher the value of c', the more deformable the soil is, since there are steeper reductions in the CP height as the blows are applied in the Mini-MCV compaction test. Thus, the soil with the highest deformability was soil 01 (c' = 2.0) indicating the behavior of a clay soil (c'>1.5 according to Nogami and Villibor (2009)), and soil 09 presented the lowest c' value showing typical behavior of non-plastic sands and silts (c'<1.0, according to Nogami and Villibor (2009)). In the permanent deformation tests, soil 09 showed the lowest deformation and soil 01 was the most notable deformation.

Regarding granulometry, we stress that other parameters associated with granulometry were also considered, but deleted from the table in order to optimize the presentation of the data. The variables in this case were: % of total sand, % of passing material in sieve no. 10, % of material retained in sieve no. 10 and percentage of passing material in no. 200.

All soils presented appreciable content of clay (between 32.28% and 50.17%) and fine sand (between

© 2022 Global Journals

18.66% and 55.85%), and low percentage of coarse material (retained in sieve no. 10 – between 0.12% and 3.76%). Regarding the consistency indexes (LL, PL and PI) it is observed that for all soils there was an increase in values with drying. The actual density values of the grains obtained were between 2.629 – 2.699 g/cm^3 which is approximate, according to Gidigasu (1976), to the quartz mineral (2.65 to 2.66 g/cm^3) and the clay mineral kaolinite (2.60 to 2.68 g/cm^3), which is coherent since most of the soils studied are composed of sand and clay fractions.

The results of maximum dry specific mass (ρ_{dmax}, g/cm^3) and optimum moisture (W_{or}, %) ranged between 1.69 - 1.97 g/cm^3 and between 10.6 - 19%, respectively, and showed some correspondence with the granulometry. Soil 03, for example, has a large percentage of fine sand in its composition, which is reflected in the compaction curve (low optimum moisture and high specific dry mass), while soil 08, with the highest percentage of silt, compared to the other soils, presented higher moisture and lower specific mass.

Six soils presented CBR values below 20%, which is, according to the paving manual of the Brazilian Department of Transport Infrastructure (DNIT, 2006), the minimum allowed for application in subbase layers, considering the empirical sizing method. The other soils presented values higher than 20%, reaching a maximum of 35% (sample 03), which according to the mentioned guidelines could be recommended for subbase layer and none of the soils would be recommended for the base layer since CBR>60% is required. Regarding expansion, six soils showed an expansion of 0% after 96 hours of immersion in water, and the other 7 samples showed low expansion values. The values of $_{mean}$ MR of all soils were above 400 MPa, considered a high value for fine-grained soils, comparable to the values of boulder soils.

As for permanent deformation, all soils presented low total deformation values (ε_p, mm) for all stress levels applied in the test. According to The National Pavement Sizing Method (MeDiNa), the sum of the contribution of all layers and subgrade to the sinking of the wheel track should be a maximum of 10 mm for Main Arterial Route, for example.

Almost half of the soils exhibit opposite behavior in relation to the parameter Ψ_2 of the expression of permanent deformation, associated with the confining stress. Soils with negative values of Ψ_2 show a reduction in permanent deformation with increased confining stress. All values of Ψ_3 are positive indicating that the variation of the deviation stress increases permanent deformation, which is expected.

The pH measurement reflects the active acidity of the soil, and the results obtained (< 5) represent soils with high acidity, which is expected for lateritic soils. Specifically in water, values between 4.4 and 4.7 were obtained indicating the presence of exchangeable aluminum (Sobral et al., 2015), which suggests possible gains in chemical stabilization processes. ΔpH is associated with the predominance of clay minerals such as kaolinite and illite, (Farias, 2012 and Camapum de Carvalho et al., 2015).

The CEC presented low values (between 2.8 – 5.74 cmol$_c$dm^{-3}) indicating predominance of 1:1 clay such as kaolinite (Gidigasu, 1976; Das, 2008 and Sobral et al., 2015). Regarding Organic Matter (OM), all soils have a low content (< 15 g.kg^{-1}) according to Prezotti (2013). For all soils, the percentage of base saturation (V, %) is considered "low" (< 50%) according to Prezotti, 2013). Aluminum saturation (S, %) was in the "low" class (< 50%) in soils 01, 02, 03 and 07, "middle" class (50% - 70%), in 04, 06, 08, 10, 11 and 13, and, "high" class (>70%) in soils 05, 09 and 12, according to Prezotti, (2013).

Table 6: Data of the characteristic curve of the studied soils (Continuation of the database in Table 5)

Soil	S_{Wo}	θ_s	θ_r	SMR	Ψ_{b1}	Ψ_{res1}	S_{res1}	Ψ_{b2}	S_b	Ψ_{res2}	S_{res2}
01	200	20,58	1,47	198,59	3,5	5,0	0,84	10000	0,82	26800	0,02
02	800	20,67	0,00	760,23	3,0	6,0	0,89	7500	0,85	22500	0,08
03	30	34,34	0,06	1071,43	2,5	16,0	0,70	8000	0,62	21500	0,02
04	30	17,64	0,00	487,86	4,0	5,5	0,74	10000	0,59	23000	0,03
05	300	14,35	0,26	434,80	3,4	6,5	0,78	8700	0,72	27000	0,03
06	50	29,31	0,99	372,62	4,8	6,5	0,84	18000	0,77	36000	0,10
07	6	23,60	1,01	430,23	6,0	9,0	0,89	14000	0,83	20000	0,05
08	100	22,26	0,00	485,99	3,8	10,0	0,91	12000	0,87	48000	0,07
09	400	13,49	0,26	1088,10	1,2	11,0	0,75	14000	0,72	30000	0,04
10	25	23,91	0,45	1527,01	2,1	20,0	0,74	15000	0,66	25000	0,04
11	40	33,60	0,17	1673,32	3,8	20,0	0,79	13500	0,71	22500	0,04
12	1000	17,20	0,00	1147,13	2,0	11,0	0,69	5500	0,66	24000	0,04
13	40	26,85	0,60	275,91	3,5	6,5	0,73	13800	0,65	30000	0,04

© 2022 Global Journals

Table 6 presents the results of the modeling of the soil-water characteristic curves. The characteristic curves of the studied soils present high suction values and air intake points, since the more weathered the soil, the higher the presented suction values are (Boszczowski and Ligocki, 2012). The shape of the curves suggests a bimodal behavior, indicating that both micropores and macropores control water inlet and outlet flows (Feuerharmel et al., 2006).

The high suction values obtained are also similar to those obtained by Marinho and Stuermer (2000) when they studied a mature residual soil of Gnaisse (45% clay), compacted in normal and modified Proctor energies, obtaining air intake values ranging between 1000 and 2000 kPa and residual suction of 15,000kPa.

b) Cluster Analysis

The dendrogram created by processing the codes in Python resulted in the hierarchical structure presented in Figure 3. Seeking to understand how the algorithm effectively grouped the data, a top-down dendrogram analysis was carried out, i.e., noting which geotechnical variable or variables may have been used as a soil divider in each Cluster. For this, three cut-off points were defined in the dendrogram, as highlighted in the figure, and the clusters created were identified. It is emphasized that the clusterization obtained went through a much more complex process than is described here, since more than fifty geotechnical attributes were considered.

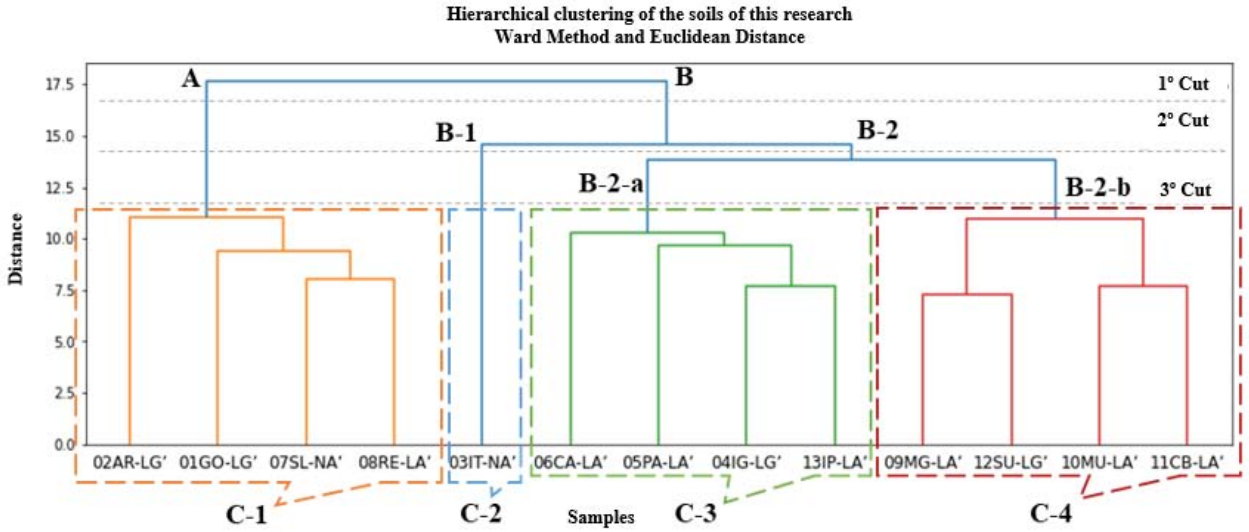


Figure 3: Analysis of the dendrogram of the hierarchical clustering of the soils of this research

Initially, in the first cut-off point, the algorithm divided the soils into two large groups (A and B), which may have been formed considering the results of the compaction test (optimum moisture and maximum dry apparent specific mass), since, as can be seen in the graph of Figure 4, there is a clear division of these groups through the compaction curves. Thus, it is observed that group A is composed of soils with higher values of optimum moisture (Wo > 16%) and lower values of ρ_{dmax} (< 1,8 g/cm³) and group B of soils with Wo < 16% and ρ_{dmax} >1,8 g/cm³), It was also noted that group A is composed entirely of USCS's ML class soils that take into account LL and PI.

© 2022 Global Journals

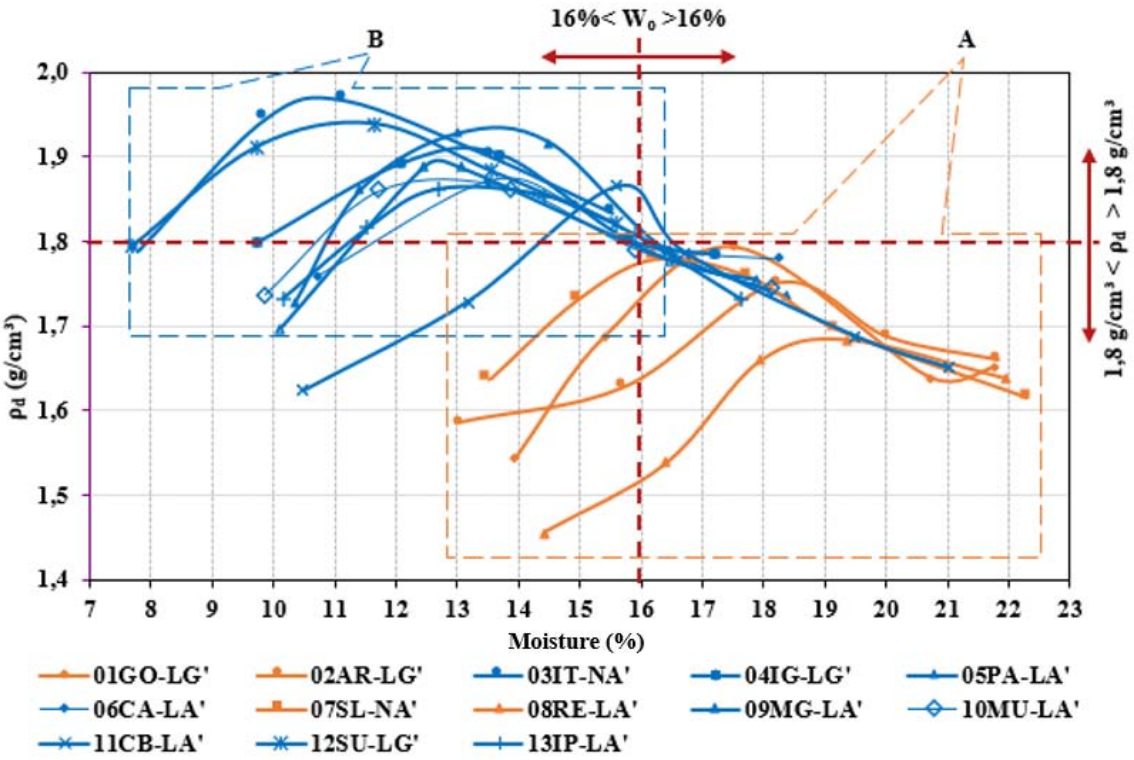


Figure 4: Division of groups A and B by the soil compaction curves of this research

In the second cut-off point, it is observed that group B is subdivided into two other groups (B-1 and B-2), while group A remains the same (showing that the soils of this group have more homogeneous characteristics than the soils of group B). This part of the clustering may have been made based on the granulometric composition, especially by the silt and fine sand fractions of the soils. The subdivision of group B into B-1 and B-2 was due to the high % of fine soil sand 03IT-NA' (55.85%), which differentiated it from the other soils of the group, as can be seen in the graph of Figure 5.

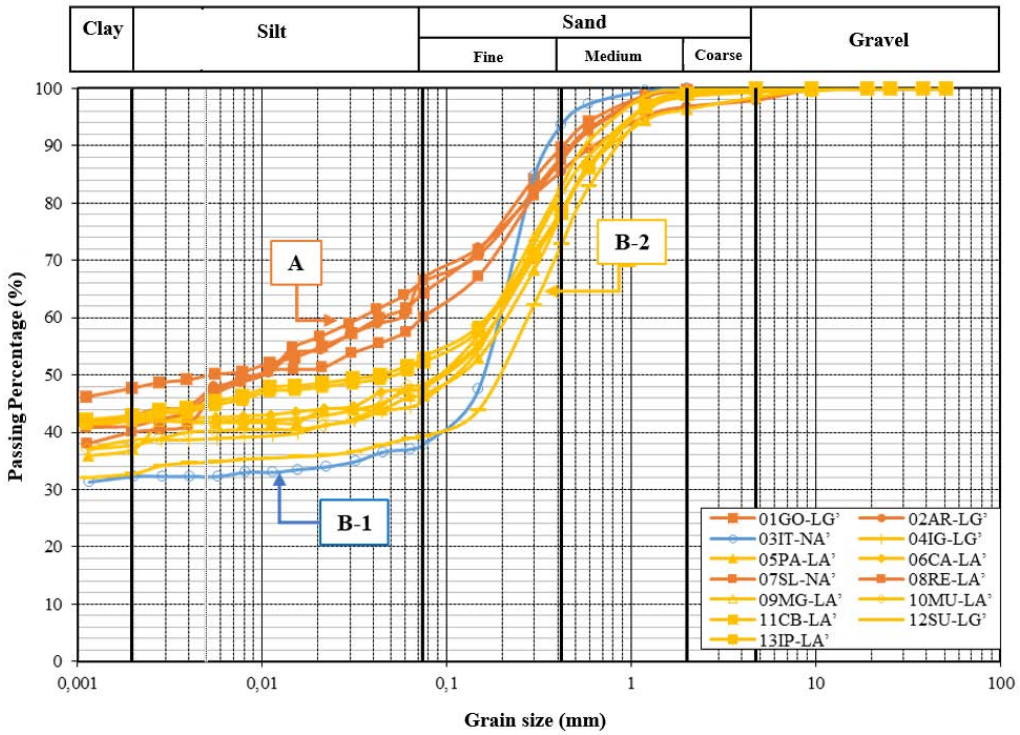


Figure 5: Division of groups A, B1 and B2 by the granulometric curves

© 2022 Global Journals

In the third and last cut-off point, group A again remained homogeneous. Because group B-1 was composed only of soil 03IT-NA', it also had no alterations (and will not have). group B-2, however, was subdivided into two subgroups (B-2-a and B-2-b), whose division may have been carried out as a function of the organic matter content found in the soils, as shown in Figure 6 (B-2-a: O.M. between 4.37 – 5.53 g/kg; B-2-b: 5.68 – 6.67 g/kg), but may also be associated with the behavior of several other variables that, despite presenting a certain "overlap" of values between groups tend, to present lower or higher results such as: ε_p - B-2-a: between 1.63 – 7.19 mm; B-2-b: 0.78 – 3.37 mm; ψ_{b1} - B-2-a: between 3.4 – 4.8 kPa; B-2-b: 1.2 – 3.8 kPa.

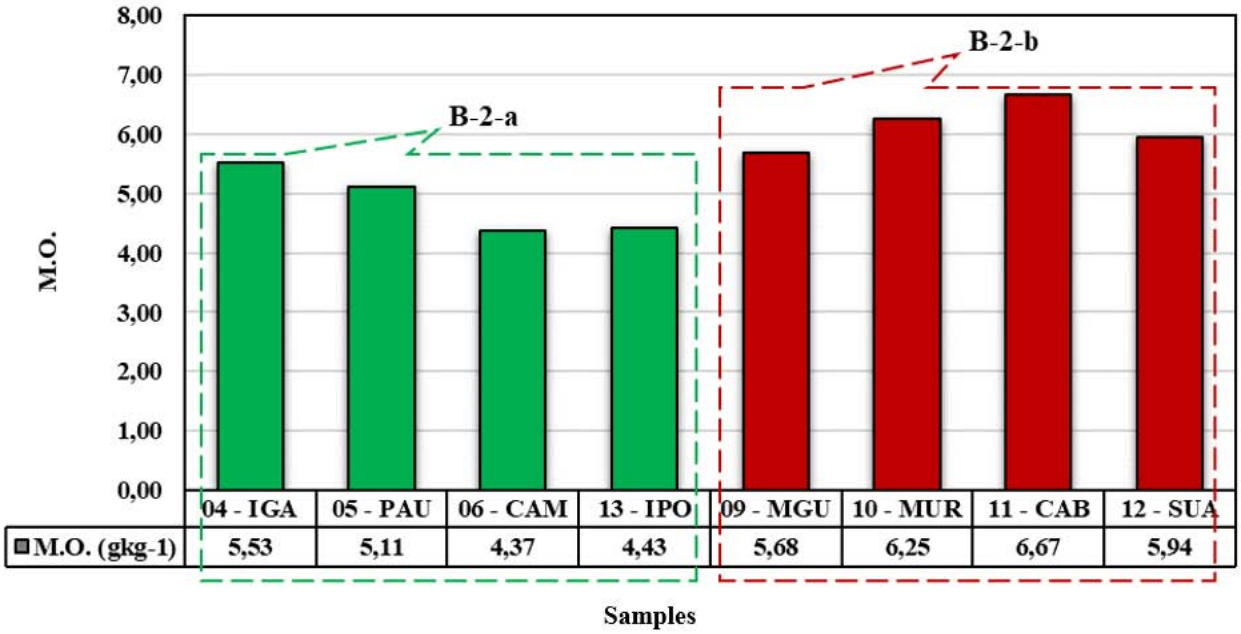


Figure 6: Division of groups B-2-a and B-2-b considering the organic matter content of the soils of this research

The final interpretation of the group allows the division of the studied soils into four different clusters (C-1 to C4) in the dendrogram, cluster 1 (C-1) composed of 4 soils of group A (01GO-LG', 02AR-LG', 07SL-NA' and 08RE-LA'), Cluster 2 (C-2) composed of the soil of group B-1 (03IT-NA'), Cluster 3 (C-3) formed by 4 soils of subgroup B-2-a (04IG-LG', 05PA-LA', 06CA-LA' and 13IP-LA') and Cluster 4 (C-4) composed of 4 soils of subgroup B-2-b (09MG-LA', 10MU-LA', 11CB-LA' and 12SULG'). From each cluster, greater similarities were identified between some pairs of soils, namely: c-1 - 07SL-NA' and 08RE-LA', C-3 - 04IG-LG' and 13IP-LA' and C-4 - 10MU-LA' and 11CB-LA'.

Soils 07SL-NA' and 08RE-LA', despite being of different MCT classes, are of identical USCS class and have similar values of liquidity limit, %clay, pH_{KCl}, δ, ε_p and S_b. Soils 04IG-LG' and 13IP-LA', though also of different MCT classes, are of the same TRB class (A-6) and have analogous values of PI, optimum moisture, K2TC, K2TD, ε_p, U'_{b1} and S_{res1}. The soils 10MU-LA' and 11CB-LA' are of the same MCT class with similar values of c', e', % silt, PI, ε_p and Ψ_{res1}.

It was also observed that the classification of aluminum saturation indicated by Prezotti (2013) may also be related to general soil grouping, since only one soil in each saturation class (low, medium and high) is distant from the formed group, which is certainly due to the other variables considered.

c) Stereo Miscroscope Image Comparison

Figures 7, 8 and 9 present the images that are most similar to the images of the 04IG-LG', 08RE-LA' and 12SU-LG' samples, respectively (a sample of each cluster obtained in the hierarchical cluster, with the exception of the cluster composed of the sample 03IT-NA because it is a group with a single element). The numbers presented next to each ranked image refers to the Euclidean distance in relation to the reference image.

The soils with color distribution most similar to soil 04IG-LG' are soils 07SL-NA', 13IP-LA', 11CB-LA', 10MU-LA' and 03IT-NA', and it is noted that, with the exception of sample 07SL-NA', the indicated soils are part of the same large hierarchical group (B) and, in addition, the 13IP-LA' soil coincides with the soil closest to 04IG-LG' in cluster C3/B-2-a. Visual similarity regarding the predominance of quartz particles is noted.

© 2022 Global Journals

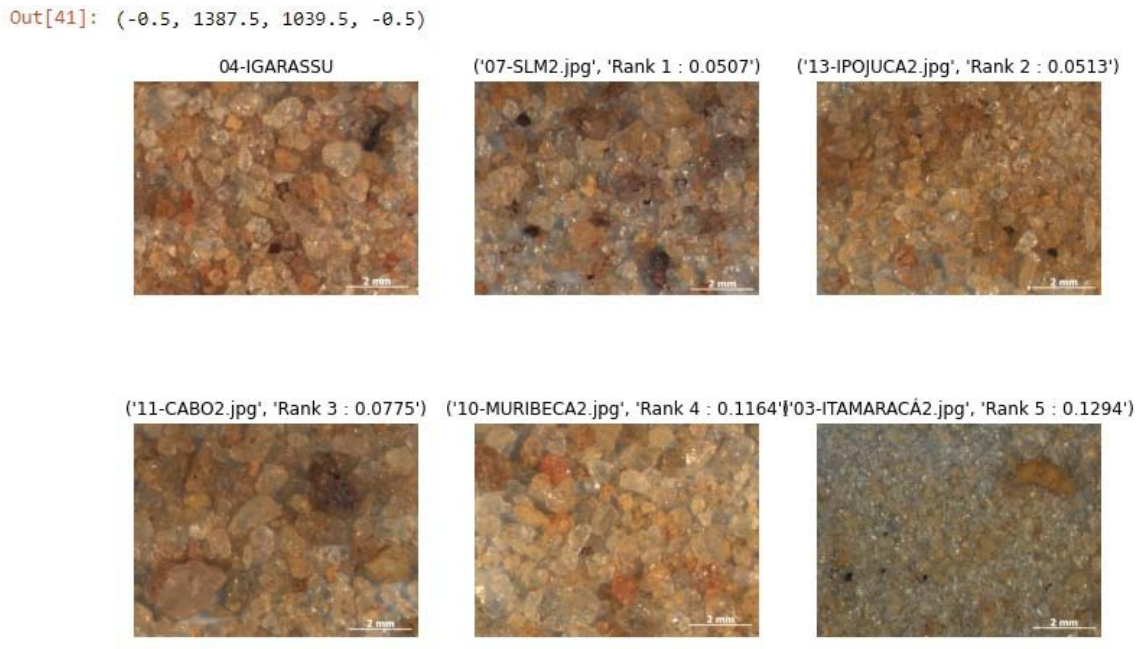


Figure 7: Images similar to sample image 04IG-LG', fine fraction

The soils with color distribution most similar to soil 08RE-LA' are soils 01GO-LG', 06CA-LA', 02AR-LG', 09MG-LA' and 12SU-LG', and it is noted that with the exception of sample 06CA-NA', the two closest soils are soils 01GO-LG' and 02AR-LG' which are part of the same large hierarchical group (A/Cluster C1). Visual similarity is also noted in relation to the presence of iron oxides (limonite).

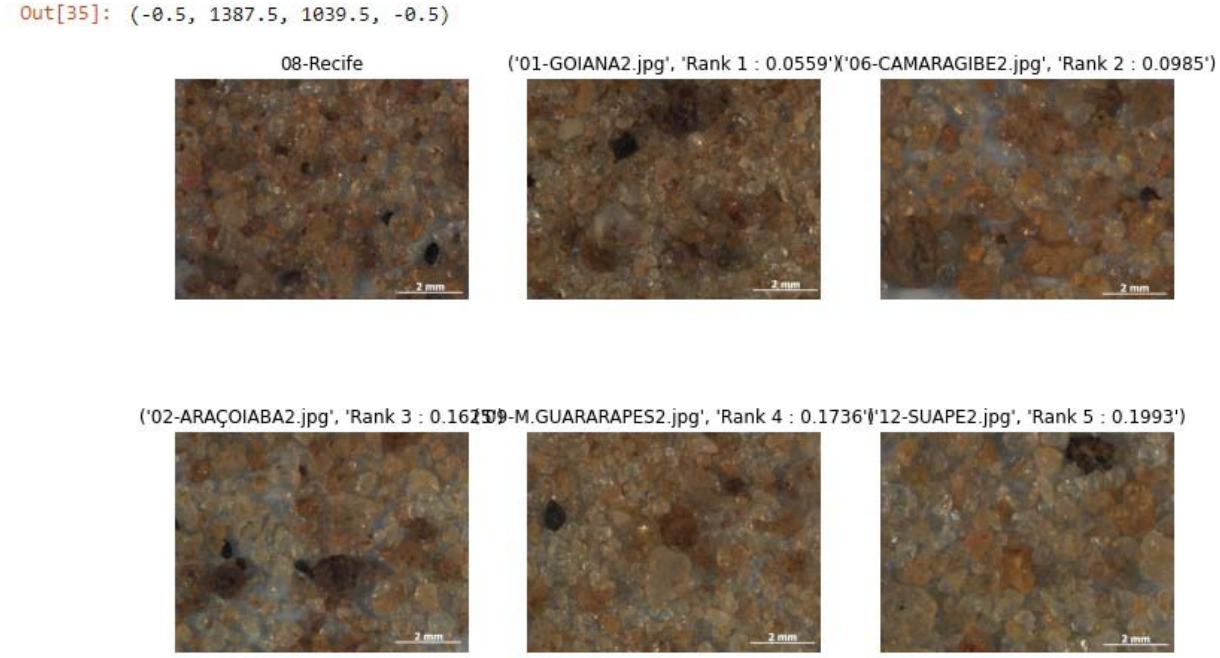


Figure 8: Images similar to sample image 08RE-LA', fine fraction

© 2022 Global Journals



Figure 9: Images similar to the 12SU-LG sample image,' fine fraction of this research

The soils with color distribution most similar to soil 12SU-LG' are soils 09MG-LA', 02AR-LG', 05PA-LA', 06CA-LA' and 01GO-LG'. It is noted that the closest soil (09MG-LA') coincides with the result of the cluster analysis, and, with the exception of soils 01GO-LG' and 02AR-LG', the other soils indicated as close are part of the same large group and subgroup (B and B-2).

In view of the obtained results, the comparison of soils through images using the techniques of data science can be considered promising, since the obtained results were near the cluster obtained with cluster analysis. The results using these two techniques, even though the same parameters were not used, showed correspondence between the mineralogical characteristics visualized in the images (iron oxides, rock fragments, quartz grains, etc.) with the results of geotechnical tests. The few variations found can be attributed to the methods themselves, since each technique used can provide different results (distance measurements, linkage methods, image descriptor, etc.).

As mentioned in the topic of Materials and Methods, for the acquisition of images, a Stereo Microscope Zeiss Discovery V8 available in the Paleontology laboratory of the Department of Geology of Federal University of Pernambuco was used. Although not trivial, it is believed that access to this type of equipment in universities is not very difficult, since it is an essential equipment in geology laboratories, biology and medicine. It can be noted that to use the Data Science tools presented here, it is not required that the images be obtained by this type of specific microscope, in fact, the important thing is that the images are of good quality and that they are obtained using minimum standards (process of obtaining samples, fraction of material, approximations, etc.) in order to obtain reliable results.

IV. Conclusion

The thirteen tropical soils of fine granulation examined in this study, originating from the northeastern region of Brazil, were characterized as materials with physical and mechanical behavior varying between clayey and sandy, having been labeled into different classes of fine soils according to the considered Classification systems (MCT, TRB and USCS). They presented high values of resilient modulus, low levels of permanent deformation, unsaturated behavior of the bimodal type in all soil-water characteristic curves, as well as chemical and mineralogical characteristics indicative of typically lateritic soils.

Multivariate analyses, such as cluster analysis applied in this article, consider three or more variables to characterize the behavior of the analyzed object, so it is understood that several geotechnical parameters were used at the same time to form the groups with the most homogeneous characteristics.

In this sense, other clusters were tested excluding, for example, the parameters of the two-dimensional models of resilient behavior (keeping only those associated with the composite model that presented the best framework for most soils), and it was observed that by excluding only the data associated with the model, or due to the deviation stress or the confining stress, there is no change in the dendrogram, however, by removing all the parameters associated with the two models, a dendrogram with another cluster

© 2022 Global Journals

structure is obtained, evidencing the association that the algorithm makes between the many variables.

Still, it was possible to notice that the mechanical characteristics of soil compaction (ρ_{dmax} and W_o) were decisive for the initial division of the groups, as well as the granulometry (mainly the percentage of silt). The chemical classification of aluminum saturation (S, %) indicated by Prezotti (2013) also showed a relationship with the clustering of the soils, indicating that the clustering considered, in fact, different types of characteristics of fine-grained tropical soils (physical, mechanical and chemical).

It was also noted that soils with mechanical clayey behavior (higher optimum moisture and lower dry apparent density) showed a more homogeneous behavior forming a cluster composed entirely of soils of the ML class (from the USCS classification, which is based on the LL and the PI) and with very similar characteristics of granulometry (percentage of silt and fine sand), in addition to the geotechnical compaction characteristics indicated.

In the group formed by soils of sandy behavior (B), the subdivision was done also by considering, in addition to the parameters considered in group A, the organic matter content, and it was noticed that the association between several other variables was also considered, since there was a tendency towards lower or higher values in some parameters, for example: in terms of total permanent deformation (ε_p) in subgroup - B-2-a: soils with ε_p larger (1.63 – 7.19 mm) than those of group B-2-b (ε_p between 0.78 – 3.37 mm) were included; the suction in the first at the air intake point associated with macropores (ψ_{b1}) tended to be slightly higher in the soils of group B-2-a (3.4 – 4.8 kPa) when compared to B-2-b (1.2 – 3.8 kPa).

The recognition of the similarity between some pairs of soils proved the validity of the hierarchical clustering technique since several variables with similar values were identified between them, even though they were of a different nature (physical, chemical and mechanical), the most recurrent being: LL, PI, %Clay, % silt, c', e', pH_{KCl}, δ, S_b, Ψ_{b1}, Ψ_{res1}, S_{res1}, W_o, K2TC, K2TD, ε_p.

The comparison of images applying the techniques of data science also corroborated, since satisfactory results were obtained, congruent with those obtained in cluster analysis, with rare exceptions. It is assumed, therefore, that there is correspondence between the mineralogical characteristics visualized in the images (iron oxides, rock fragments, quartz grains, etc.) with the results of geotechnical tests.

Finally, it is concluded that the application of cluster analysis by hierarchical method, as well as the comparison of microscopic images, using the tools of Data Science, showed useful techniques and tools for the cluster analysis of fine-grained tropical soils since it portrayed the similarity of behavior of different soils considering several geotechnical aspects.

Acknowledgements

The authors thank the Foundation for Science and Technology of Pernambuco (FACEPE) for the financial support (doctoral scholarship to the first author) and incentive to carry out this research. The research was carried out through the INCT REAGEO Project through the partnership signed between GEGEP (Geotechnical Engineering Group of Disasters and Plains) of THE PPGEC of UFPE (Federal University of Pernambuco) and the Geotechnics Laboratory of COPPE/UFRJ (Federal University of Rio de Janeiro).

References Références Referencias

1. Boszczowski, R. B., & Ligocki, L. P. (2012). Chapter 8: Características Geotécnicas dos solos residuais de Curitiba e RMC. In Twin Cities – Solos das Regiões Metropolitanas de São Paulo e Curitiba. D'Livros. ISBN: 978-85-86438-40-0. [In Portuguese].
2. Camapum de Carvalho, J., de Rezende, L. R., Cardoso, F. B. F., Lucena, L. C. F. L., Guimarães, R. C., & Valencia, Y. G. (2015). Tropical soils for highway construction: Peculiarities and considerations. Transportation Geotechnics, 5, 3–19. https://doi.org/10.1016/j.trgeo.2015.10.004.
3. Chen, C. Härdle, W. K. Unwin, A. 2007. *Handbook of Data Visualization.* Springer-Verlag Berlin Heidelberg. 936 p. DOI: 10.1007/978-3-540-33037-0.
4. Chesworth, Ward et al. 2008. "Tropical soils". *Encyclopedia of Soil Science*, [s.l.], p.793-803. Springer Netherlands. http://dx.doi.org/10.1007/978-1-4020-3995-9_607.
5. Coutinho, R. Q., & Sousa, M. A. S. (2021). Analysis of the applicability of USCS, TRB and MCT classification systems to the tropical soils of pernambuco, Brazil, for use in road paving. In Advances in transportation geotechnics IV. Lecture Notes in Civil Engineering, Vol. 164. Springer. https://doi.org/10.1007/978-3-030-77230-7_29.
6. Coutinho, R. Q., Silva, M. M., Santos, A. N. D., & Lacerda, W. A. (2019). Geotechnical characterization and failure mechanism of landslide in granite residual soil. Journal of Geotechnical and Geoenvironmental Engineering, 145(8), 05019004. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002052.
7. CPRM, Companhia de Pesquisa de Recursos Minerais. 2014. *Geodiversidade do estado de Pernambuco*. ISBN 978-85-7499-141-2. Recife. [In Portuguese].
8. Dalla Roza, A. E. Motta, L. M. G. Classificação MCT com relação ao comportamento resiliente e deformação permanente em solos do Mato Grosso.

© 2022 Global Journals

In: XIX Congresso Brasileiro de Mecânica dos Solos e Engenharia Geotécnica – COBRAMSEG. Anais... Bahia, 2018. [In Portuguese].
9. Das, B. M. (2008). Advanced soil mechanics (3rd ed.). Taylor & Francis.
10. Farias, W. M. (2012). Processos Evolutivos de Intemperismo Químico e Sua Ação no Comportamento Hidromecânico de Solos do Planalto Central [PhD thesis] University of Brasilia (UnB). [In Portuguese]. 263 p. https://repositorio.unb.br/handle/ 10482/11902. [In Portuguese].
11. Feuerharmel, C., Gehling, W. Y. Y., & Bica, A. V. D. (2006). The use of filter-paper and suction-plate methods for determining the soil-water characteristic curve of undisturbed colluvium soils. Geotechnical Testing Journal, 29(5), 419–425. https://doi.org/10.1520/GTJ14004.
12. Forsyth, D. 2018. *Probability and Statistics for Computer Science*. Springer International Publishing AG.
13. Frank, I. E. and Todeschini, R. 1994. *The data analysis handbook*. Elsevier.
14. Gidigasu, M. D. (1976). Laterite soil engineering: Pedogenesis and engineering principles. Elsevier. ISBN 0-444-41283-2.
15. Gitirana, G. F. N., & Fredlund, D. G. (2004). Soil-water characteristic curve equation with independent properties. Journal of Geotechnical and Geoenvironmental Engineering, 130(2), 209–212. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:2(209).
16. Grus, J. 2016. Data Science do Zero. Tradução de: *Data Science from Scratch: First Principles with Python.* traduzido por Welington Nascimento. Rio de Janeiro: Alta Books.
17. Guimarães, A. C. R. (2009). Um Método Mecanístico Empírico para Previsão da Deformação Permanente em Solos Tropicais Constituintes de Pavimentos [PhD thesis]. Federal University of Rio de Janeiro (UFRJ). [In Portuguese].
18. Guimarães, A. C. R., Motta, L. M. G., & Castro, C. D. (2019). Permanent deformation parameters of fine – grained tropical soils. Road Materials and Pavement Design, 20(7), 1664–1681. https://doi.org/10.1080/14680629.2018.1473283.
19. Guimarães, A. C. R.; Silva Filho, J. C. and Castro, C. D. 2021. "Contribution to the use of alternative material in heavy haul railway sub-ballast layer". *Transportation Geotechnics*, Volume 30, 100524, ISSN 2214-3912, https://doi.org/10.1016/j.trgeo. 2021.100524.
20. Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. 2009. *Análise multivariada de dados*. Bookman editora.
21. Hardle, W. K. Simar, L. 2015. *Applied Multivariate Statistical Analysis*. Fourth Edition. Springer: Verlag Berlin Heidelberg.
22. Harris, C.R., Millman, K. J., van der Walt, S.J. *et al.* 2020. "Array programming with NumPy". *Nature* 585, 357–362. DOI: https://doi.org/10.1038/s41586-020-2649-2.
23. Harris, C.R., Millman, K. J., van der Walt, S.J. *et al.* 2020. "Array programming with NumPy". *Nature* 585, 357–362. DOI: https://doi.org/10.1038/s41586-020-2649-2.
24. Hastie, T.; Tibshirani, R. and Friedman, J. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
25. Lima, C. D. A. Motta, L. M. G. Guimarães, A. C. R. 2017. "Estudo da deformação permanente de britas granito-gnaisse para uso em base e sub-base de pavimentos". *Revista Transportes.* Volume 25, Número 2. DOI: 10.14295/transportes.v25i2.1262. [In Portuguese].
26. Lima, C. D., Motta, L. M. G. da, Aragão, F., & Guimarães, A. C. R. (2020). Mechanical characterization of fine-grained lateritic soils for mechanistic-empirical flexible pavement design. Journal of Testing and Evaluation. https://doi.org/10.1520/JTE20180890.
27. Marinho, Fernando A. M.; Stuermer, Mônica M. 2000. "The Influence of the Compaction Energy on the SWCC of a Residual Soil". *Advances In Unsaturated Geotechnics*, [s.l.], p. 125-141, 24 jul. American Society of Civil Engineers. http://dx.doi.org/10.1061/40510(287)8.
28. Mayssa Alves da Silva Sousa, Roberto Quental Coutinho & Laura Maria Goretti da Motta (2021): Analysis of the unsaturated behaviour of compacted lateritic fine-grained tropical soils for use in transport infrastructure, Road Materials and Pavement Design, DOI: https://doi.org/10.1080/14680629. 2021.2009008.
29. MEDINA, J.; PREUSSLER, E. Características Resilientes de Solos Em Estudos de Pavimentos. SOLOS E ROCHAS. (REV. BRAS. DE GEOTECNIA)., ABMS/ABGE, v. 3, n.1, p. 3-26, 1980. [In Portuguese].
30. Medina, J., & Motta, L. M. G. (2015). Mecânica dos Pavimentos. 2nd ed. Editora Interciência. [In Portuguese].
31. Metz, J. 2006. *Interpretação de cluster gerados por algoritmos de clustering hierárquico*. Dissertação de Mestrado. Universidade de São Paulo, USP. São Paulo, Brasil. [In Portuguese].
32. Millman, K. J. and M. Aivazis. 2011. "Python for Scientists and Engineers". *Computing in Science & Engineering*. vol. 13, no. 2, pp. 9-12, March-April. DOI: 10.1109/MCSE.2011.36.
33. Nogami, J. S., & Villibor, D. F. (1991). Use of lateritic fine-grained soils in road pavement base courses. Geotechnical and Geological Engineering, 9(3–4), 167–182. https://doi.org/10.1007/BF00881739.

© 2022 Global Journals

34. Nogami, J. S., & Villibor, D. F. (1995). Pavimentação de Baixo Custo com Solos Lateríticos. Editora Villibor. [In Portuguese].
35. Pedregosa, F. 2011. "Scikit-learn: Machine learning in Python". *The Journal of machine Learning research,* v. 12, p. 2825-2830.
36. Prezotti, L. C. 2013. *Guia de interpretação de análise de solo e foliar*. 104 p. Incaper: Espírito Santo, Brasil. [In Portuguese].
37. Sobral, L. F. Barreto, M. C. V. Silva, A. J. Anjos, J. L. 2015. *Guia prático para interpretação de resultados de análises de solos*. Embrapa Tabuleiros Costeiros: Aracaju, Brasil. [In Portuguese].
38. Sousa, M. A. S. (2016). Análise geotécnica de solos tropicais de ocorrência ao longo da Estrada de Ferro Carajás para uso como camadas de pavimento rodoviário ferroviário [Master thesis]. Military Institute of Engineering (IME)]. [In Portuguese]. http://transportes.ime.eb.br/DISSERTA%C3%87%C3%95ES/2016%20MAYSSA%20ALVES%20DA%20SILVA%20SOUSA.pdf [In Portuguese].
39. Svenson, M. (1980). Ensaios triaxiais dinâmicos de Solos argilosos. Federal University of Rio de Janeiro (UFRJ). [In Portuguese].
40. VanderPlas, Jake. 2016. Python Data Science Handbook: Essential Tools for Working with Data. Estados Unidos: O'Reilly Media.
41. Villibor, D. F., & Nogami, J. S. (2009). Pavimentos econômicos: Tecnologia do uso dos solos finos lateríticos. Arte & Ciência. 291 p. [In Portuguese].
42. Virtanen, P., Gommers, R., Oliphant, T.E. *et al.* 2020. "SciPy 1.0: fundamental algorithms for scientific computing in Python." *Nature Methods.* 17, 261–272. DOI: https:// doi.org/10.1038/s41592-019-0686-2.

© 2022 Global Journals