scientific reports

Check for updates

OPEN

Statistical evaluation of testing conditions on the saturated hydraulic conductivity of Brazilian lateritic soils using artificial intelligence approaches

Weber Anselmo dos Ramos Souza[1], Sávio Aparecido dos Santos Pereira[2,3], Thiago Augusto Mendes[2]✉, Rafaella Fonseca Costa[4], Gilson de Farias Neves Gitirana Junior[3] & Juan Félix Rodríguez Rebolledo[5]

The saturated hydraulic conductivity, k_{sat}, is a crucial variable to describe the hydromechanical behavior of soils. The value of k_{sat} of lateritic soils that are typically found in tropical regions is highly affected by the soil's structure, void ratio, and fine particle aggregation. As a result, the determination of k_{sat} in the field or in the laboratory is complex and involves greater variability, depending on the type of test and on the spatial location of sampling. This paper presents a study of k_{sat} values of lateritic soils, analyzing them using Statistic, Multilayer Perceptron Artificial Neural Networks (ANN) and Decision Trees (CHAID). This study aims to support decision-making regarding the type of test and depth chosen for sampling in laterite soils and understanding the factors influencing the permeability of such soils. An extensive literature review on the k_{sat} values of lateritic soils was performed, providing data for the establishment of a database comprise of 722 registries. According to agronomic and geotechnical soil classifications, the Brazilian lateritic soils presents a "moderate" hydraulic conductivity. A significant variation of permeability values along the depth was identified, particularly for depths between 0.1 and 0.2 m. Regarding the importance of testing variables, the ANN indicated a high dependency on the type of test. The decision tree divided field test and laboratory test automatically, inferring the relevance of the type of test to the determination of k_{sat}.

The hydraulic conductivity of the soil (k) is defined, by Darcy's law, as the relationship between the percolation rate of water volume per unit of total area and the hydraulic head gradient. When k reaches its maximum value, it is called saturated hydraulic conductivity of the soil (k_{sat}), a property that is dependent on the soil particle size distribution, particle morphology, pore continuity, particle orientation, volume of pores, among other factors[1–3]. The testing methodology for determining k_{sat}, whether in-situ or laboratory, also influences its value.

The dependence of k_{sat} on these several factors turns this into a complex parameter with significant variability, reaching variations of over 200%[3–6]. In this context, understanding how k_{sat} is affected by soil characteristics, the type of test performed and the sample depth becomes essential for determining this parameter[7–9].

In lateritic soils, k_{sat} becomes particularly affected by the soil structure. Moreover, this soil, typical of tropical regions from Africa, South America, and Southeast Asia[10,11], are commonly formed from the weathering of rocks subjected to the high temperatures and humid climates typical of these regions[10,12,13], and are characterized by often having a high void ratio, high clay content, and significant presence of iron and aluminum oxides and hydroxides, resulting in the aggregation of these fine particles[10,14]. Particle aggregation form macro and micropores in the soil, resulting in a bimodal pore-size distribution. This structure particularity of the soil pores offers

[1]Institut de Recherche en Mines et en Environnement, Université du Québec, Abitibi-Témiscamingue, Rouyn-Noranda, Québec, Canada. [2]Federal Institute of Education, Science and Technology of Goias (IFG), Aparecida de Goiânia, Brazil. [3]School of Civil and Environmental Engineering, Federal University of Goias (UFG), Goiânia, Brazil. [4]Department of Civil, Construction, and Environmental Engineering, North Caroline State University, Raleigh, North Caroline, USA. [5]Technology College, University of Brasilia (UnB), Brasília, Brazil. ✉email: thiago.mendes@ifg.edu.br

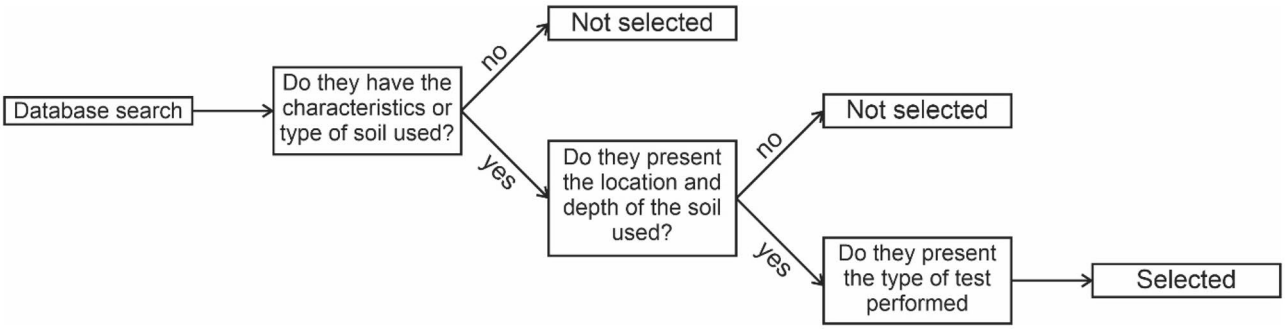


Figure 1. Criteria for selecting publications.

preferential paths for water percolation, which can generate large variations in the values of k_{sat} when comparing different samples obtained from the same location. As a result, understanding the permeability phenomena in this soil type becomes challenging[13].

The physical determination of k_{sat} may employ field or laboratory tests. In field tests, the soil does not undergo significant deformations in its structure, allowing a better understanding of the permeability phenomenon. However, it is difficult to control the boundary conditions of field tests. The main field tests used for determining k_{sat} are: Guelph permeater, slug test, pump test and concentric ring or double ring infiltrometer. The main laboratory tests available are: permeameters with constant and falling head, and triaxial tests.

Field and laboratory tests are costly and require considerable times[15]. For his reason, indirect estimation methods have been developed, such as theoretical equations[3,16], pedotransfer functions[15,17,18], and machine learning methods[19–21]. Machine learning methods, which can have either a regression or classification character, can be an essential tool for assessing the sensitivity of variables influewnced by complex relationships, as is the case for k_{sat}.

This paper evaluates of the main factor influencing k_{sat} values of Brazilian lateritic soils using statistics, multilayer perceptron artificial neural networks (ANN), and a CHAID-type decision tree. The analyses presented herein contribute to the understanding of hydraulic conductivity in Brazilian lateritic soils aind in the establishemtn of adequate methodologies for the determination of representative k_{sat} values, according to the sample depth and test type.

Materials and methods

Data collection. First of all, a wide literature review was carried out on scientific articles that deal with the determination of the saturated hydraulic conductivity (k_{sat}) of Brazilian lateritic soils, especially from the Midwest region. Test results on soils with the presence of roots were disregarded. The data collected included various types of equipment, testing methodologies, and variable soil sample depth. The following databases were used: (a) Scopus, (b) Web of Science, (c) ASCE and (d) Google Scholar, emphasizing the latter due to the ease, accessibility, and practicality in searching and obtaining journals, dissertations, and theses.

Search strings were established to identify publications presenting the measurement of the hydraulic conductivity of lateritic soils, considering various testing methodologies, equipment, and depths. Initially, the geographic location was not specified since lateritic soils are present in several Brazilian locations. There was no limitation on the publication period for the scientific articles researched. Combination of words for publications written in Portuguese were: ("permeabilidade" AND "ensaio" AND "laterítico") OR ("condutividade hidráulica" AND "ensaio" AND "laterítico") OR ("condutividade hidráulica saturada" AND "ensaio" AND "laterítico"). The corresponding combinations of words for publications written in English were ("permeability" AND "test" AND "lateritic") OR ("hydraulic conductivity" AND "test" AND "lateritic") OR ("saturated hydraulic conductivity" AND "test" AND "lateritic").

Altogether, 6414 scientific articles were found: 60 from Scopus, 10 from Web of Science, 154 from ASCE, and 6190 from Google Scholar. Because of the large number of articles found, these were screened based on the rejection criteria shown in the flowchart presented in Fig. 1.

Using the collection of k_{sat} data obtained in the literature review, a database was built to support statistical and computational analyses, aiming to understand better the influence of the type of test, method or equipment and depth of sample. Machine learning methods, including Artificial Neural Networks (ANN) and decision trees, were used as described later.

Materials. Data used in this paper comprise exclusively lateritic soils. According to Fortes and Merighi[22], lateritic soils' colors are yellow and red because of aluminum hydroxides and ferric hydrates. The Unified Soil Classification System (USCS) is generally considered not suitable for tropical soils. In fact, lateritic materials are better characterized and classified using the MCT (Miniature, Compacted, Tropical) classification system. In terms of location, lateritic soils are typically found in the regions indicated by the Charman[23] map (Fig. 2).

The magnitude of the k_{sat} of tropical soils can be classified according to the flow rate ranges established by Ferreira (1999) apud Freire et al.[24], and presented in Table 1. It could be noted that the values shown in Table 1 are for agronomic studies, and were initially presented in meters day^{-1}, being converted to meters seconds^{-1} in this paper. A permeability classification for use in civil works was not found in the literature. However, a classification of permeability values by type of soil was published by Das and Sobhan[25], as shown in Table 2.

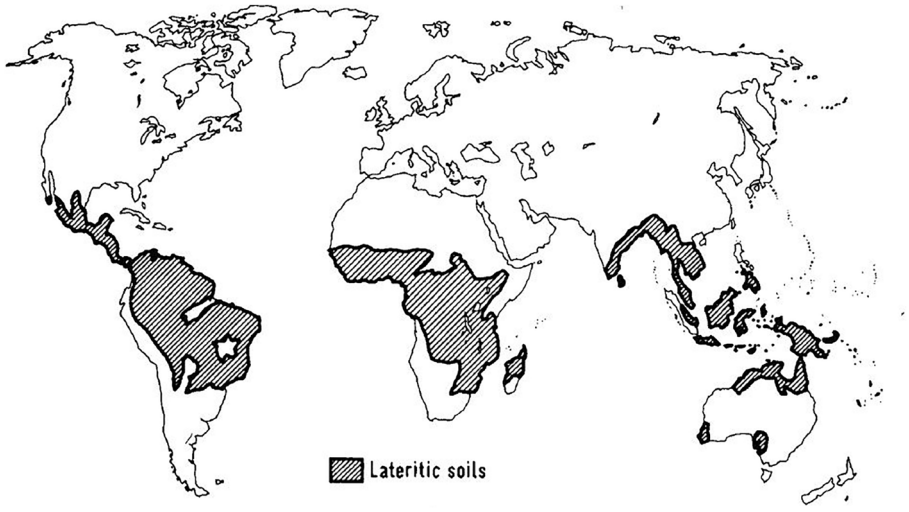


Figure 2. Regions with the occurrence of lateritic soils (Charman[23]).

Classes	k_{sat} (m s^{-1})
Very fast	$>6.94\times10^{-5}$
Fast	3.47×10^{-5}–6.94×10^{-5}
Moderately fast	1.74×10^{-5}–3.47×10^{-5}
Moderately	5.56×10^{-6}–1.74×10^{-5}
Moderately slow	1.39×10^{-6}–5.56×10^{-6}
Slow	3.47×10^{-7}–1.39×10^{-6}
Very slow	$<3.47\times10^{-7}$

Table 1. Saturated hydraulic conductivity classes established by Ferreira (1999) apud Freire et al.[24].

Classes	k_{sat} (m s^{-1})
Clear gravel	1×10^{0}–1×10^{-2}
Coarse sand	1×10^{-2}–1×10^{-4}
Fine sand	1×10^{-4}–1×10^{-5}
Silty clay	1×10^{-5}–1×10^{-7}
Clay	$<1\times10^{-8}$

Table 2. Saturated hydraulic conductivity classes presented by Das and Sobhan[25].

Regarding the types of tests on laterite soils, the two main laboratory apparatuses used to determine k_{sat} are the constant load and the falling head permeameters. These tests are standardized in Brazil through technical standards NBR 13,292[26] and NBR 14,545[27], respectively. Undisturbed and remolded specimens can be used in these tests, with the constant head test being commonly adopted for granular soils, while the falling head test is aimed at clay soils with relatively low permeability.

The main approached for determining k_{sat} in the field are the Guelph test, the slug test, the pump test, and the concentric rings test. The Guelph apparatus is commonly used, being practical, easy to perform, having a low cost, and offering a quick means of determination of k_{sat}. The Guelph apparatus is generally used for small depths, up to 75 cm[28,29]. The slug test is a field test that consists of inserting a cylinder into the soil and monitoring the flow of water.

The pump test is a more complex procedure that uses a suction pump to assess variations in water flow in the soil. The concentric ring method consists of placing two rings of different diameters on the soil surface. Water is added to the two circles, where an initial reading of the water height is performed and readings are taken at predetermined time intervals, evaluating the height variation in the inner circle to calculate the hydraulic conductivity. This method does not have many difficulties in its execution and results in more homogeneous values. Unfortunately, the concentric ring method offers overestimated values of k_{sat}, due to the imposed hydraulic head[30].

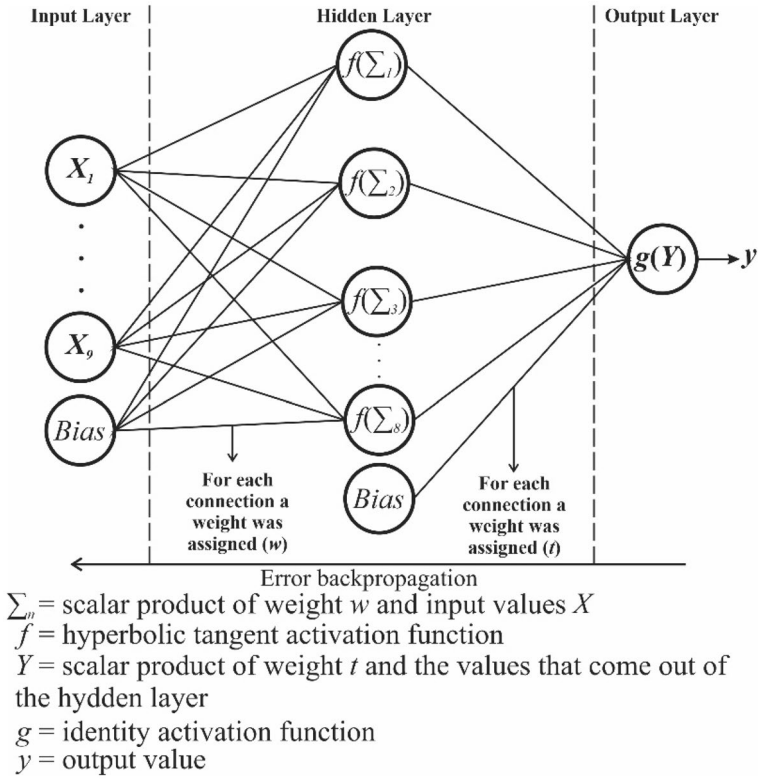


Figure 3. Diagram of the ANN adopted.

Description of the ANN used. The software used for building the ANN was IBM SPSS Statistics[31], adopting a multilayer perceptron network[32–34]. The multilayer perceptron network is a commonly adopted supervised learning method among the various ANN methodologies developed. The approach requires the availability of training data, used to adapt the model. This network is commonly used to evaluate databases, images, and other types of data[35–38], and consists of a set of layers subdivided into: input, hidden, and output layers. The input layer consists of the data region used for training, while the output layer returns the desired parameter. The hidden layer is an intermediate layer and aims to connect the input values with the output values. The connections between each layer are made through weights, which are values assigned initially at random, and later adjusted by the ANN during training. Weights are assigned from the input layer to the hidden layer and from the hidden layer to the output layer.

The values that feed the hidden layer come from the scalar product between the assigned weights and the input layer values, being applied through a mathematical function called activation function, which has the purpose of linearizing the data[39]. The most used activation functions are the ReLU and the hyperbolic tangent[34]. The final value that comes out of the hidden layer is the scalar product applied to the activation function arriving at the output layer[39]. The resulting value that leaves the output layer is the value predicted by the ANN[39].

To minimize the errors computed during training, a backpropagation process was used: the application of the descending gradient, an interactive method of non-linear optimization[33,34]. The descending gradient method requires two critical parameters: the learning rate and the momentum. The learning rate determines the learning speed of the algorithm. In general, values between 0 and 1 are adopted. Momentum is a parameter that gives the network stability, allowing for rapid convergence. Like the learning rate, its values vary between 0 and 1.

Another parameter determined during training is the Bias, a unitary component used to compensate for random weight assignments[34]. Its incorporation into an ANN is essential, as it allows the translation of the scalar product between the components of a layer and their respective weights, preventing the ANN from assigning greater weight to a component of a specific layer due to the lack of freedom of movement of the scalar product.

Figure 3 outlines the ANN model used, based on the concepts, parameters and methodology described herein. As activation functions, hyperbolic tangent and identity were used for the input and the output, respectively. In the modeling of the ANN, a single hidden layer was used. The ANN method with a multilayer perceptron network used in this paper offers also a quantitative assessment of the importance of the variables used in the determination of k_{sat}, through the assigned weights.

A fraction of 69.3% of the data was used for network training, 20.1% for testing, and 10.6% for model validation (holdout). In the validation, cross-validation was adopted, and the training was carried out in batches. Cross-validation is characterized by dividing the data into a specific integer value n, selecting for each n a percentage of data for training and another for testing. The process is repeated for each division n and the model with the best fit is adopted. The initial learning rate was 0.4 and the moment was 0.9.

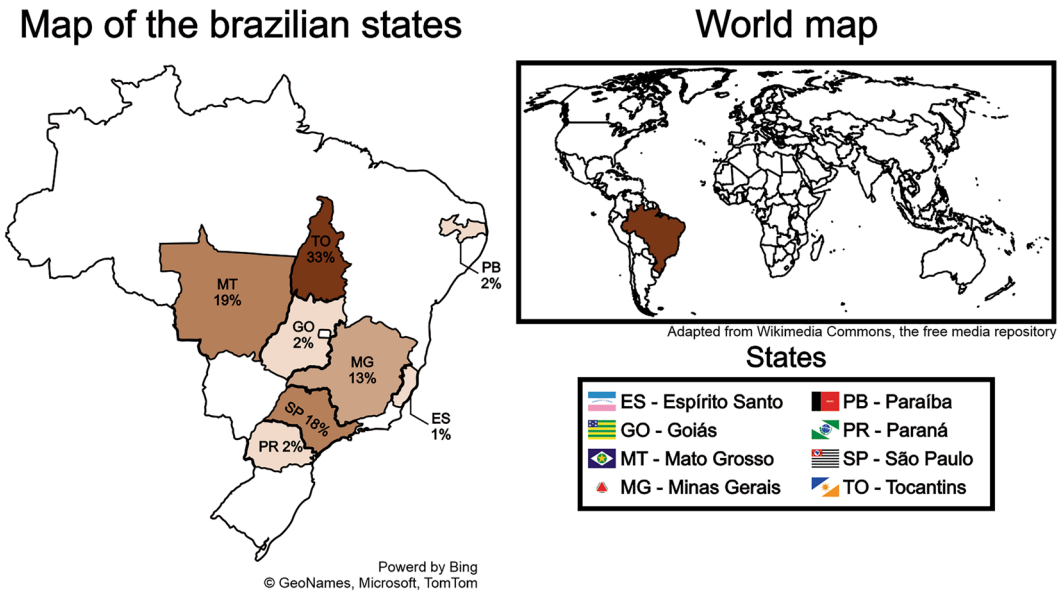


Figure 4. Geographic distribution of selected k_{sat} data.

Chis-square automatic interaction detector (CHAID) decision tree. According to Biggs et al.[40], decision tree and classification techniques are powerful tools used for dividing analysis data into homogeneous groups. One of these methods is the automatic interaction using chi-square (CHAID), implemented through an algorithm developed by Kass[41] and improved by Biggs et al.[40]. The population subdivision criteria must meet the selected statistical significance, maximum depth of the tree and minimum number of cases (parent node and child node). If the criteria not reached, the values are not divided concerning the studied variable[42].

According to Kass[41], CHAID comprises the following steps: (a) best partition for each predictor; (b) the predictors are compared and the best one is chosen; (c) the data are subdivided according to the predictors and the division criteria; (d) each subgroup is independently analyzed to produce other subdivisions.

The advantages of CHAID are its straightforward interpretation and reading, little computational time and the possibility of having multiple divisions in a node[42,43]. Regarding the disadvantages, the method requires a large amount of data to achieve adequate results[44]. This method was used herein to evaluate the influence of the different variables involved (types of tests, method or equipment and sample depth) in the determination of k_{sat}, proposing a class of the importance of these variables through the generated model (CHAID), is also used for validating the results of ANN modeling.

Results and discussions

Overview of selected publications. A total of 6414 scientific papers, theses and dissertations were found, but only 18 were selected, 7 of which were doctoral theses[45–51], 6 master's dissertations[52–57], 1 monography[58] and only 4 scientific papers[14,59–61]. In addition to the documents searched in the databases, the k_{sat} data presented in the Brazilian Agricultural Research Corporation (EMBRAPA) bulletin, Filizola et al.[9], was also included. The geographic distribution of the k_{sat} data selected is shown in Fig. 4.

The selected publications presented results obtained using the constant head, the falling head, the Guelph, the double ring, the flexible wall, and the field-falling permeameters. Some studies also presented data obtained from infiltration well, and well pumping tests. The depths of the tests or samples varied from 0.00 to 3.00 m. Figure 5 shows the distribution of k_{sat} registries by each method. The most common laboratory permeability test was the constant head, and the most common field test was the Guelph test. According to Elmashad and Ata[62], the most common field test is the ring infiltrometer, but not the Guelph test.

Analysis of collected data. The main statistical parameters describing the collected soil permeability dataset are presented in Tables 3, with k_{sat} values in their original form and in the natural logarithmic scale. Table 3 shows higher coefficient of variation (COV) values, while logarithmic ones show a significant reduction. This reduction is associated with the scaling effect of the logarithm operator.

The k_{sat} values of the Brazilian lateritic soils surveyed were classified according to the flow rate ranges shown in Tables 1 and 2. The corresponding ranges, in terms of infiltration speed are shown along with the box diagrams presented in Figs. 6 and 7. The box plot limits are presented for $p = 0.05$.

The values of the natural logarithm of k_{sat} (Fig. 6a) were plotted in the box diagrams, considering that these values follow the lognormal distribution, as suggested by Gitirana and Fredlund[63], and the original k_{sat} values (Fig. 6b). It is important to note that the original k_{sat} values have an asymmetric distribution. The use of permeability values in their original form, as shown in Fig. 6b, leads to the identification of outliers only above the

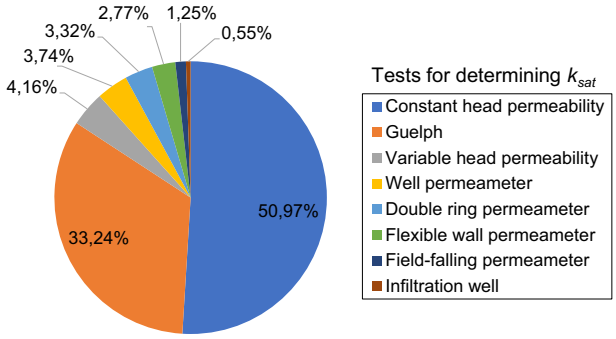


Figure 5. Distribution of the methods used in the collected data.

Group	Min	Max	Median	Mean	Std. Dev	COV %	N
Original scale							
Constant head	4.00E−09	1.22E−03	1.13E−05	1.56E−05	6.32E−05	404.8%	368
Falling head	6.56E−09	7.40E−05	2.08E−06	6.90E−06	1.47E−05	212.6%	30
Guelph	2.11E−08	1.50E−03	8.58E−06	7.17E−05	1.77E−04	246.9%	240
Double ring permeameter	4.17E−07	8.30E−05	8.89E−06	1.69E−05	2.25E−05	133.3%	24
Flexible wall permeameter	2.70E−09	1.42E−05	6.08E−06	6.59E−06	5.94E−06	90.0%	20
Field-falling permeameter	2.64E−07	1.33E−05	4.73E−06	5.79E−06	5.19E−06	89.6%	9
Infiltration well	1.60E−04	1.52E−03	7.77E−04	8.09E−04	7.38E−04	91.3%	4
Well permeameter	1.00E−06	1.00E−04	1.00E−05	3.09E−05	3.85E−05	124.7%	27
Natural logarithmic scale (absolute values)							
Constant head	6.71	19.34	11.39	11.53	0.92	8.0%	368
Falling head	9.51	18.84	13.14	13.29	1.98	14.9%	30
Guelph	6.50	17.67	11.67	11.59	2.19	18.9%	240
Double ring permeameter	9.40	14.69	11.63	11.91	1.56	13.1%	24
Flexible wall permeameter	11.16	19.73	12.01	13.06	2.35	18.0%	20
Field-falling permeameter	11.23	15.15	12.26	12.73	1.48	11.6%	9
Infiltration well	6.49	8.74	7.60	7.61	1.23	16.2%	4
Well permeameter	9.21	13.82	11.51	11.11	1.23	11.1%	27

Table 3. Data statistics for k_{sat} values and absolute values of $ln(k_{sat})$. *COV* Coefficient of variation, *N* number of samples.

superior range limit. These evidences indicate that the statistical interpretation of the natural logarithm of k_{sat} may provide more intuitive and useful information.

It can be inferred from Fig. 6a that the values between the first quartile and the third quartile, comprising 50% of the dataset, in addition to the median and average data, were in the "moderately slow", "moderate" and "moderate fast" classification classes. According to the criteria proposed by Ferreira (1999) apud Freire et al.[24], this indicates that the infiltration capacity of the studied lateritic soils is high. Although there is a wide variety of textural characteristics among Brazilian oxisols, previous studies have not shown a clear relationship between these characteristics and hydraulic properties[59]. The dataset presents numerous outliers with respect to the depth data, both above and below the range limits, reflecting its wide sample variability. This fact can be explained by the different types of tests used to obtain the k_{sat} values.

Figure 6a shows that most of the k_{sat} data surveyed fall within the range classified as "very fast", mainly for low depths, up to 60 cm, then moving to the "fast" class for depths between 0.6 and 1.20 m, and increasing again to the "very fast" class for greater depths of the soil profile i.e., higher than 1.20 m).

In Fig. 7, the k_{sat} values are presented along with depth ranges and separated according to the ranges shown in Table 2. In Fig. 7a, the permeability values are presented in the natural logarithm, while in Fig. 7b the values are presented on their original scale. Again, it is possible to notice that, in Fig. 7b, the diagram is skewed, since the hydraulic conductivity data do not follow the normal distribution and are asymmetric[63]. A significant portion of the values is found in the range of clay with silt and fine sand, showing the moderate infiltration capacity of lateritic soils presented in Fig. 6a.

To assess the influence that the type of equipment, instrument or test method may have on the determination of k_{sat}, three approaches were adopted: (a) 3D graphing; (b) Artificial Neural Networks (ANN) and (c) CHAID Decision Tree. Figure 8 shows all the k_{sat} data, with an average k_{sat} equal to 4.5×10^{-6} m s^{-1}. The dataset

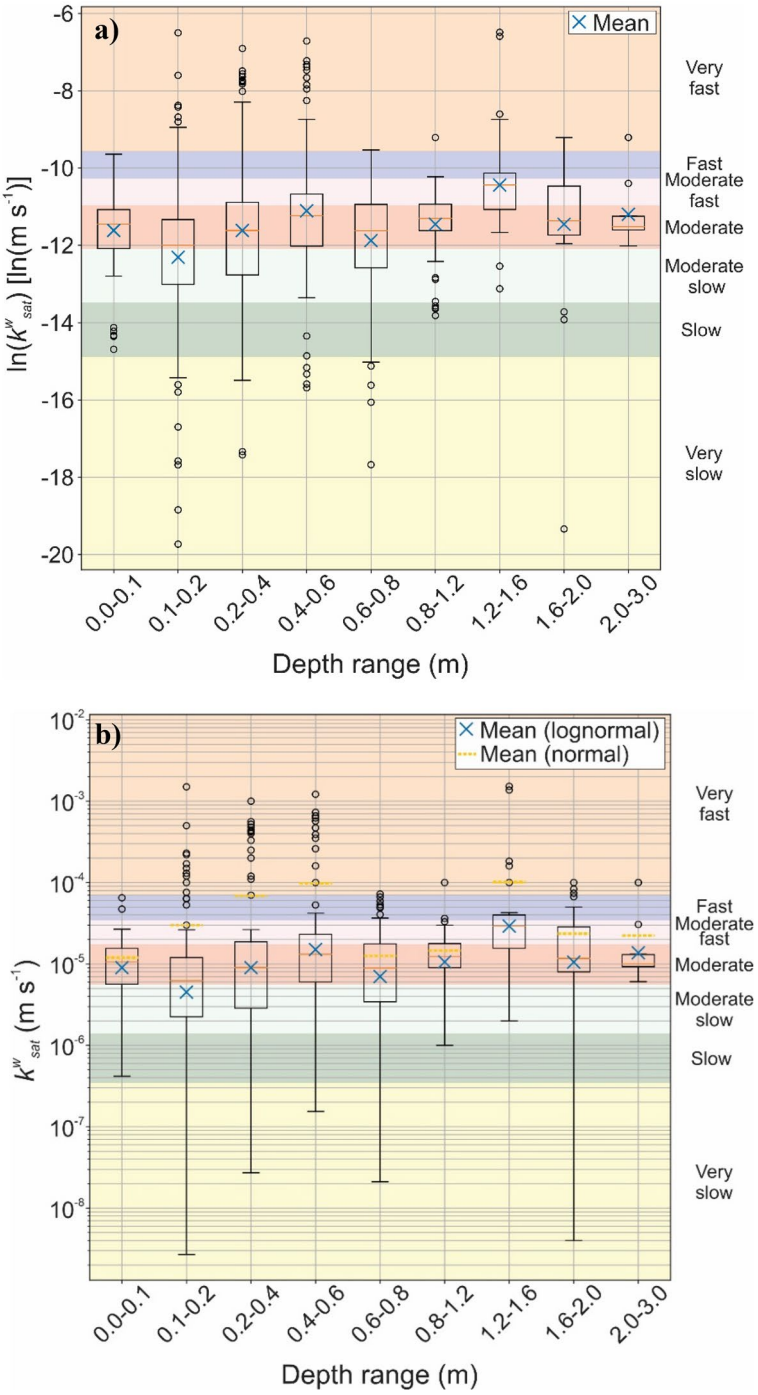


Figure 6. Distribution of saturated hydraulic conductivity of lateritic soils by depth range, classified according to Table 1: (**a**) ln k_{sat}; (**b**) k_{sat}.

is separated according to the type of permeameter (constant head, falling head and Guelph) and infiltrometer. This average value is similar to that obtained using triaxial permeameter tests, as presented by Mendes[64], Vaz[65], and Araújo[54,66], all for lateritic soil from the city of Goiânia, state of Goiás, Brazil.

It could be noticed that the average k_{sat} values are overestimated when combining all depths (Fig. 8). This is, once again, a results of the skewed distribution, as presented in Fig. 6a.

Figure 9 presents the procedure for assessing the influence of testing conditions (i.e., test type and testing depth) using the ANN. In Fig. 9, each box represents the layer variables that return the value of the next layer. The thickness of the connections between each box represents the modulus of the weight assigned to that value in determining the value of the next layer. The greater the thickness of the connection, the greater the value of the modulus of the weight of that element. In addition, the weight module signals are represented in gray for

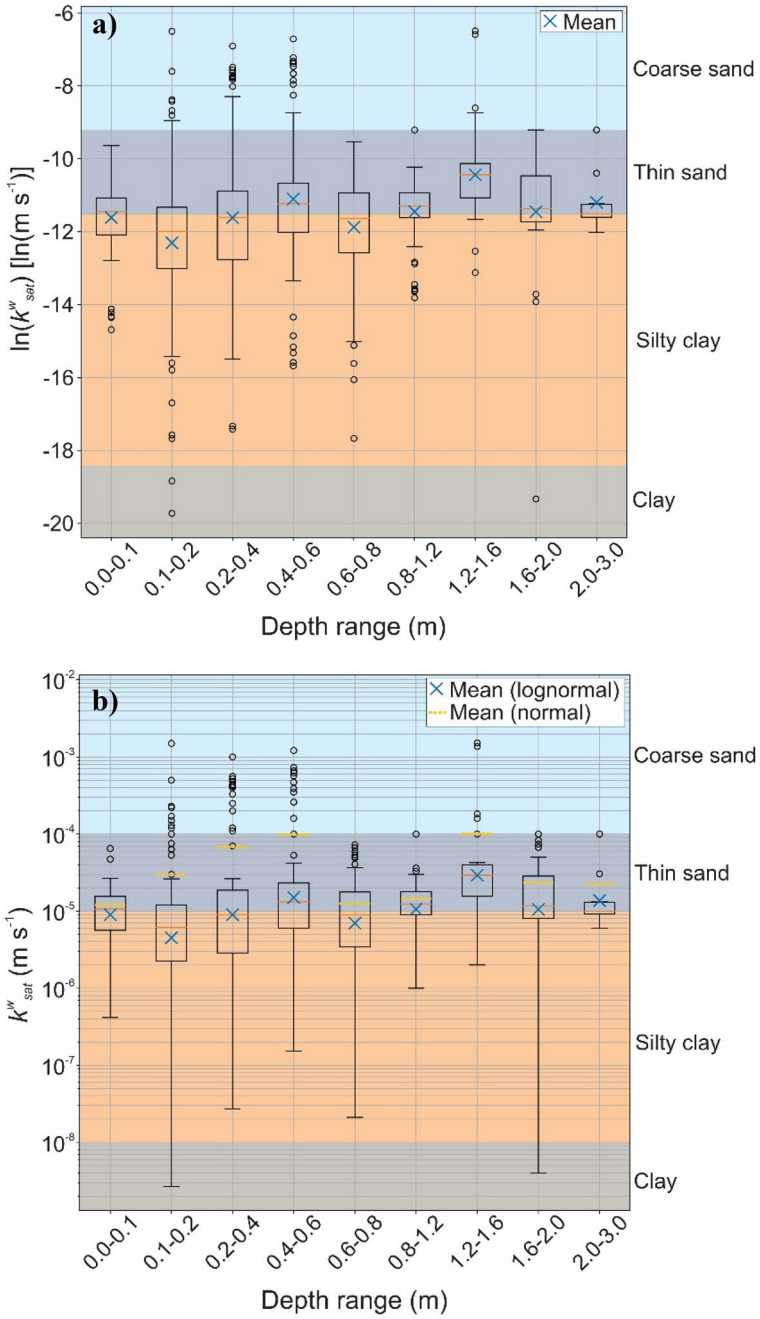


Figure 7. Distribution of the saturated hydraulic conductivity of lateritic soils by depth range, classified according to Table 2: (**a**) ln k_{sat}; (**b**) k_{sat}.

positive values and blue for negative values. The weight signals are interpreted as being directly proportional in determining the value of the next layer, when positive, and inversely proportional when negative.

A sensitivity analysis was made based on the training and testing samples to determine each predictor's importance in the neural network. These values represent the relative importance for the main predictor. The testing method had a 82.2% relative influence, whereas the testing depth has a 17.8% relative influence. Figure 9 indicates that the type of equipment, instrument and test performed has a stronger connection with the hidden layer of the ANN and, consequently, more significant importance in the determination of k_{sat} than the sample depth.

Comparing the results obtained using the ANN (Fig. 9), a decision tree of the CHAID type (Fig. 10) was developed using a division criterion node relative with a minimum of 60 tests, and child node with a minimum of 30 trials. The CHAID decision tree has three classification levels: the upper level (Node 0), determined by the k_{sat} values; the intermediate level (Nodes 1–7), showing the depths of sampling/testing and the last level (Nodes 8–11), with the types of test used for determining k_{sat}.

The levels of the decision tree indicate the hierarchy or degrees of relative importance of the parameters and variables evaluated by the algorithm, the first level being the k_{sat} values. This level was subdivided into seven

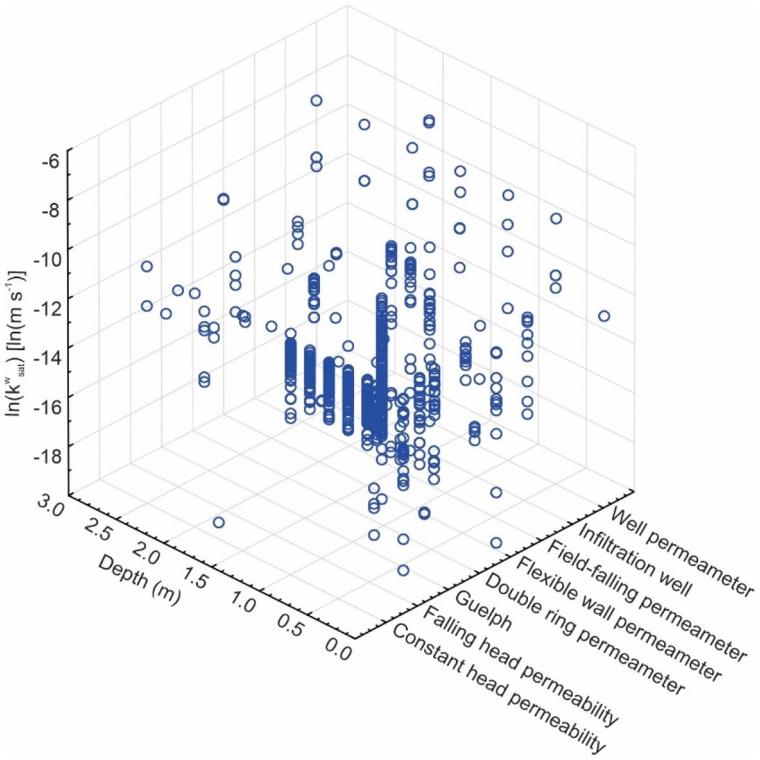


Figure 8. Distribution of the saturated hydraulic conductivity of the lateritic soil by depth range and type of test by different authors, considering all the researched data.

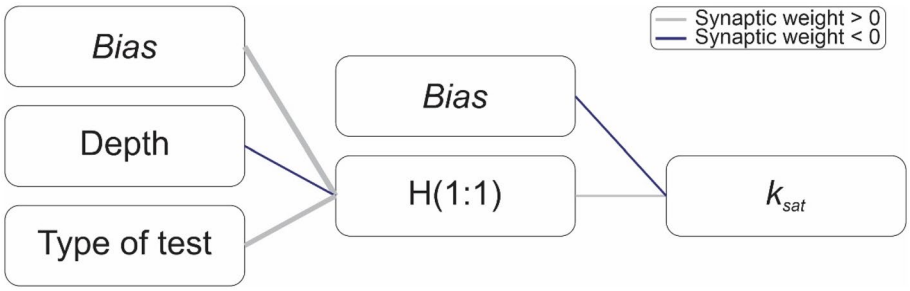


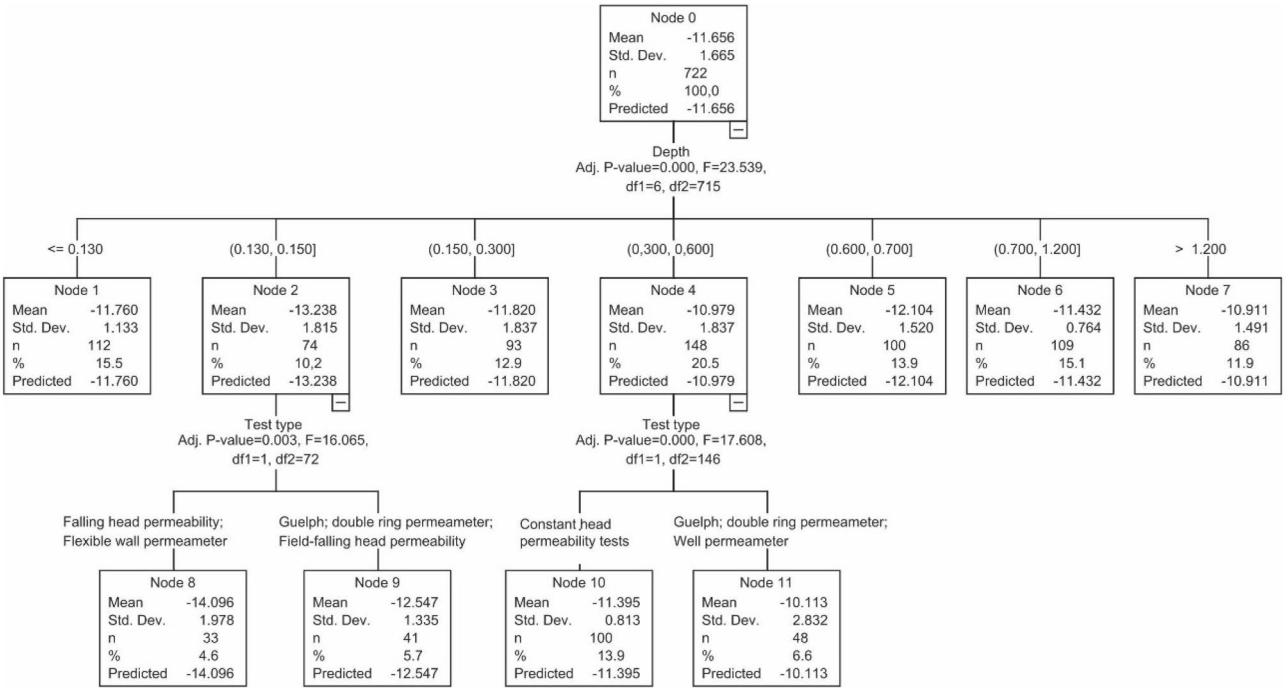
Figure 9. ANN approach for the evaluation of the hydraulic conductivity data of the lateritic soils.

branches or nodes (Nodes 1–7), with the statistical parameters of each node and its significance being indicated in each node box (Fig. 10).

For Nodes 2 and 4, that is, the most superficial depths and the most suitable for hydrological and agronomic studies (0.13–0.15 and 0.3–0.6 m), there was a division of the k_{sat} values as the representativeness of the testing method (Nodes 8–11). Thus, the algorithm classified for each node the type of test (field and laboratory) most likely recommended for the determination of k_{sat}, according to the testing depths, indicating the values that are closer to the mean and standard deviation.

Figure 10 indicates that the permeability values obtained from laboratory and field tests have characteristics in common between them. This finding may allow the optimization of testing programs and reduce costs and result in testing time savings due to the similarity between the methods. Furthermore, it was possible to observe that the k_{sat} values obtained from laboratory tests are lower than those obtained from field tests. Some authors have reported this finding previously, such as Gribb et al.[67], Elfaki[68], and Elhakim[69]. For the other nodes (other depths evaluated), there was no significant difference in k_{sat} obtained in the field or in the laboratory, as perceived by the CHAID decision tree generator algorithm.

In general, the results demonstrate a significant relevance of the type of test and sample depth to the measured k_{sat} values of Brazilian lateritic soils. The statistical assessment of the testing approach is essential for the analysis of the results obtained in testing programs. A high variability in the k_{sat} values was observed both considering the type of test and the sample depth. COV values of 247% were observed for the Guelph test and 405% for the constant head. Along the depth a variation between 10^{-2} and 10^{-9} m s^{-1} was observed for the k_{sat} values.



where: F the F test of significance; df_1 e df_2 the number of subsets -1 e n-1, respectively.

Figure 10. CHAID decision tree for hydraulic conductivity values of Brazilian lateritic soils surveyed.

Considering the average values of k_{sat}, it was possible to verify that, despite its variability, the average values were similar when comparing determinations using the same type of test and the same depth. Moreover, the constant head permeameter determinations and testings on samples obtained between 0.7 and 1.2 m in depth showed the lowest standard deviations. As a result, these testings specifications (i.e., constant head on specimens between 0.7 and 1.2 m) should result in more consistent permeability data.

Conclusions

According to the results and analyses presented herein, the the following conclusions may be drawn from this study:

(1) Most saturated hydraulic conductivity data published for Brazilian lateritic soils corresponds to laboratory tests using constant and falling head permeameters (around 65%) and Guelph field tests (around 35%), These tests where carried out for depths between 10 cm and 3 m. The obtained values of k_{sat} are usually classified as "moderate" for agronomic purposes and correspond to the behavior of sands, from a geotechnical engineering perspective.
(2) The artificial neural networks (ANN) and the CHAID decision tree proved to be efficient tools to support the selection of testing methodology, the depth of execution of the tests for the determination of k_{sat}, and to raise awareness regarding difficulties in interpreting results obtained for Brazilian lateritic soils. For instance, the depths of soil profiles between 0.13 to 0.6 m showed the highest standard deviation values. Therefore, it is necessary to provide carefull interpretations of testing results obtained from this depth range.
(3) The ANN showed that the type of test has more influence on the value of k_{sat} than the sampling depth (about 4.6 times superior significance). It is important to note that the soil condition (e.g., landfills, cuts, pre-densification) may also play a role in the value of k_{sat}, requiring further studies.
(4) It could be noted that the CHAID decision tree indicated that it is possible to separated field test and laboratory test in the same data sample. In general, the results obtained with the CHAID decision tree indicated lower values of k_{sat} obtained by laboratory tests than those performed in the field. Moreover, the constant head permeability and the range of depth between 0.7 and 1.2 m. showed the lowest standard deviation values. Therefore, this type of test and this interval of depth was suggested for lateritic soils.

Finally, it is important to emphasize that the results presented herein do not replace traditional testing programs. These results may provide preliminary estimations of k_{sat}. The reported statistical values may aid the design of testing programs, allowing a better understanding of variability of testing results.

Received: 22 December 2021; Accepted: 21 November 2022
Published online: 27 November 2022

References

1. Sch, A. E., Grez, R. & Ramírez, C. L. conductividad hidráulica en fase saturada como herramienta para el diagnóstico de la estructura del suelo. *Agro sur.* **25**(1), 51–56. https://doi.org/10.4206/agrosur.1997.v25n1-06 (1997).
2. Mesquita, M. G. B. & Moraes, S. O. A dependência entre a condutividade hidráulica saturada e atributos físicos do solo. *Ciência Rural.* **34**(3), 963–969. https://doi.org/10.1590/S0103-84782004000300052 (2004).
3. Ren, X. *et al.* A relation of hydraulic conductivity—Void ratio for soils based on Kozeny–Carman equation. *Eng. Geol.* **213**, 89–97. https://doi.org/10.1016/j.enggeo.2016.08.017 (2016).
4. Lumb, P. The variability of natural soils. *Can. Geotech. J.* **3**(2), 74–97. https://doi.org/10.1139/t66-009 (1966).
5. Uzielli, M. Statistical analysis of geotechnical data. In *Geotechnical and Geophysical Site Characterization* (eds. Huang, A. & Mayne, P. W.) 173–193 (2008).
6. Nagy, V. *et al.* Continuous field soil moisture content mapping by means of apparent electrical conductivity (ECa) measurement. *J. Hydrol. Hydromech.* **61**(4), 305–312. https://doi.org/10.2478/johh-2013-0039 (2013).
7. Risse, L. M., Nearing, M. A. & Zhang, X. C. Variability in Green-Ampt effective hydraulic conductivity under fallow conditions. *J. Hydrol.* **169**(1–4), 1–24. https://doi.org/10.1016/0022-1694(94)02676-3 (1995).
8. Gupta, R., Rudra, R. & Dkkinson, T. Modeling infiltration with varying hydraulic conductivity under simulated rainfall conditions. *JAWRA J. Am. Water Resour. Assoc.* **34**(2), 279–287. https://doi.org/10.1111/j.1752-1688.1998.tb04134.x (1998).
9. Filizola, H. F., Fontana, A., Viana, J. H. M., Luiz, A. J. B. & Souza, M. D. *Diagnóstico de atributos físico-hídricos dos solos de textura arenosa em áreas de intensificação agrícola no bioma Cerrado* 1–74 (Embrapa Meio Ambiente—Boletim de Pesquisa e Desenvolvimento (INFOTECA-E), 2019).
10. Amu, O. O., Adetayo, O. A. & Alabi, O. S. Modification of cement stabilized structural lateritic pulverized snail shell. *Acta Technica Corviniensis Bull. Eng.* **12**(4), 63–68 (2019).
11. Mahapatra, S. & Jha, M. K. On the estimation of hydraulic conductivity of layered vadose zones with limited data availability. *J. Earth Syst. Sci.* **128**(3), 1–17. https://doi.org/10.1007/s12040-019-1101-1 (2019).
12. Chalermyanont, T., Arrykul, S. & Charoenthaisong, N. Potential use of lateritic and marine clay soils as landfill liners to retain heavy metals. *Waste Manag.* **29**(1), 117–127. https://doi.org/10.1016/j.wasman.2008.03.010 (2009).
13. Jha, M. K., Mahapatra, S., Mohan, C. & Pohshna, C. Infiltration characteristics of lateritic vadose zones: Field experiments and modeling. *Soil Tillage Res.* **187**, 219–234. https://doi.org/10.1016/j.still.2018.12.007 (2019).
14. Rodriguez, T. T., Weiss, L. A., Teixeira, R. S. & Branco, C. J. M. C. Permeabilidade de solo laterítico por diferentes métodos. *Semina Ciências Exatas e Tecnológicas.* **36**(2), 17–32. https://doi.org/10.5433/1679-0375.2015v36n2p17 (2015).
15. Picciafuoco, T. *et al.* A pedotransfer function for field-scale saturated hydraulic conductivity of a small watershed. *Vadose Zone J.* **18**(1), 1–15. https://doi.org/10.2136/vzj2019.02.0018 (2019).
16. den Putte, A. V. *et al.* Estimating the parameters of the Green-Ampt infiltration equation from rainfall simulation data: Why simpler is better. *J. Hydrol.* **476**, 332–344. https://doi.org/10.1016/j.jhydrol.2012.10.051 (2013).
17. Xu, C. *et al.* Enhancing pedotransfer functions (PTFs) using soil spectral reflectance data for estimating saturated hydraulic conductivity in southwestern China. *CATENA* **158**, 350–356. https://doi.org/10.1016/j.catena.2017.07.014 (2017).
18. Zhang, Y. & Schaap, M. G. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. *J. Hydrol.* **575**, 1011–1030. https://doi.org/10.1016/j.jhydrol.2019.05.058 (2019).
19. Erzin, Y., Gumaste, S. D., Gupta, A. K. & Singh, D. N. Artificial neural network (ANN) models for determining hydraulic conductivity of compacted fine-grained soils. *Can. Geotech. J.* **46**(8), 955–968. https://doi.org/10.1139/T09-035 (2009).
20. Kashani, M. H., Ghorbani, M. A., Shahabi, M., Naganna, S. R. & Diop, L. Multiple AI model integration strategy—application to saturated hydraulic conductivity prediction from easily available soil properties. *Soil Tillage Res.* **196**, 104449. https://doi.org/10.1016/j.still.2019.104449 (2020).
21. Gupta, S., Lehmann, P., Bonetti, S., Papritz, A. & Or, D. Global prediction of soil saturated hydraulic conductivity using random forest in a covariate-based geotransfer function (CoGTF) framework. *J. Adv. Model. Earth Syst.* **13**(4), e2020MS002242. https://doi.org/10.1029/2020MS002242 (2021).
22. Fortes, R. M. & Merighi, J. V. The use of MCT methodology for rapid classification of tropical soils in Brazil. *IJP Int. J. Pavements.* **2**(3), 1–13 (2003).
23. Charman, J. H. Laterite in road pavements. In *Construction Industry and Information Association* (CIRIA, 1988).
24. Freire, M. B. G. S. *et al.* Condutividade hidráulica de solos de Pernambuco em resposta à condutividade elétrica e RAS da água de irrigação. *Revista Brasileira de Engenharia Agrícola e Ambiental.* **7**(1), 45–52. https://doi.org/10.1590/S1415-43662003000100008 (2003).
25. Das, B. M. & Sobhan, K. *Fundamentos de engenharia geotécnica* (Thomson Learning, 2017).
26. ABNT-Associação Brasileira de Normas Técnicas. *NBR 13,292: Solo-Determinação do coeficiente de permeabilidade de solos granulares a carga constante* (2021).
27. ABNT-Associação Brasileira de Normas Técnicas. *NBR 14,545: Solo–Determinação do coeficiente de permeabilidade de solos argilosos a carga variável* (2021).
28. Soto, M. A. A., Chang, K. H. & Vilar, O. M. Análise do método do Permeâmetro de Guelph na determinação da condutividade hidráulica saturada. *Águas Subterrâneas* https://doi.org/10.14295/ras.v23i1.17004 (2009).
29. Ragab, R. & Cooper, J. D. Variability of unsaturated zone water transport parameters: Implications for hydrological modelling. 1. In situ measurements. *J. Hydrol.* **148**(1–4), 109–131. https://doi.org/10.1016/0022-1694(93)90255-8 (1993).
30. Soto, M. A. A. & Kiang, C. H. Avaliação da condutividade hidráulica em dois usos do solo na região central do Brasil. *Revista Brasileira de Ciências Ambientais (Online).* **47**, 1–11. https://doi.org/10.5327/Z2176-947820180169 (2018).
31. IBM. *Guia rápido do IBM SPSS Statistics 25. Manual de instruções* (2019).
32. McCulloch, W. S. & Pitts, W. A. Logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**(4), 115–133. https://doi.org/10.1007/BF02478259 (1943).
33. Portugal, M. S. & Fernandes, L. G. L. Redes neurais artificiais e previsão de séries econômicas: uma introdução. *Nova Economia.* **6**(1), 51–73 (1996).
34. Rauber, T. W. *Redes neurais artificiais* (2005).
35. Perelmuter, G., Carrera, E. V. C., Vellasco, M. & Pacheco, M. A. Reconhecimento de imagens bidimensionais utilizando Redes Neurais Artificiais. In *Anais do VIII Sibgrapi* (ed. Lotufo, R. A. & Mascarenhas, D. A.) 197–203 (1995).
36. Osório, F. S. & Bittencourt, J. R. Sistemas Inteligentes baseados em redes neurais artificiais aplicados ao processamento de imagens. In *I Workshop de inteligência artificial* (2000).
37. Nääs, I. A., Campos, L. S. L., Baracho, M. S. & Tolon, Y. B. Uso de redes neurais artificiais na identificação de vocalização de suínos. *Engenharia Agrícola.* **28**(2), 204–216. https://doi.org/10.1590/S0100-69162008000200001 (2008).
38. Gorgens, E. B., Leite, H. G., Santos, H. N. & Gleriani, J. N. Estimação do volume de árvores utilizando redes neurais artificiais. *Revista Árvore.* **33**(6), 1141–1147. https://doi.org/10.1590/S0100-67622009000600016 (2009).
39. Zhang, Q., Song, J. & Nie, X. Application of neural network models to rock mechanics and rock engineering. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **28**(6), 535–540. https://doi.org/10.1016/0148-9062(91)91130-J (1991).
40. Biggs, D., de Ville, B. & Suen, E. A method of choosing multiway partitions for classification and decision trees. *J. Appl. Stat.* **18**(1), 49–62. https://doi.org/10.1080/02664769100000005 (1991).

41. Kass, G. V. An exploratory technique for investigating large quantities of categorical data. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **29**(2), 119–127. https://doi.org/10.2307/2986296 (1980).
42. IBM. *SPSS I Decision trees 21* (2012).
43. Hoare, R. Using CHAID for classification problems. In *New Zealand Statistical Association Conference* (2004).
44. Nisbet, R., Elder, J. & Miner, G. *Handbook of Statistical Analysis and Data Mining Applications* (Elsevier, 2009).
45. Meaulo, F. J. *Caracterização geológica, hidrogeológica e o mapeamento da vulnerabilidade natural à poluição dos aquíferos, na escala 1: 25.000, das áreas urbana e de expansão do município de Araraquara-SP* (Universidade Estadual Paulista, São Paulo, 2007).
46. Morais, F. *Estudo dos fatores pedogeomorfológicos intervenientes na infiltração em zonas de recarga no Complexo Metamórfico Bação, Minas Gerais* (Universidade Federal de Ouro Preto, 2007).
47. Soares, A. M. S. *A dinâmica hidrológica na bacia do alto curso do rio Uberabinha-Minas Gerais* (Universidade Federal de Uberlândia, 2008).
48. Moreira, C. A. *Geofísica aplicada no monitoramento de área de disposição de resíduos sólidos domiciliares* (Universidade Estadual Paulista, 2009).
49. Silva, J. P. *Estruturas de infiltração com utilização de materiais alternativos no controle de alagamentos, inundações e prevenção de processos erosivos* (Universidade de Brasília, 2012).
50. Rosa, C. R. *Cartografia geotécnica da área costeira no município do Conde (PB): caracterização morfopedológica e processos de urbanização* (Universidade de Brasília, 2017).
51. Pierozan, R. C. *Estudo da resistência de interface de tiras metálicas e poliméricas em diferentes solos* (Universidade de Brasília, 2018).
52. Silva, J. P. *Estudos preliminares para implantação de trincheiras de infiltração* (Universidade de Brasília, 2007).
53. Freitas, S. M. A. C. *Quantificação da infiltração e da recarga de aquíferos do Alto Rio das Velhas, Minas Gerais* (Universidade Federal de Ouro Preto, 2010).
54. Araújo, A. G. *Análise do desempenho de poços de infiltração na cidade de Goiânia, Goiás* (Universidade Federal de Goiás, 2013).
55. Sírio, D. L. N. *Monitoramento e modelagem da recarga freática em técnica de drenagem compensatória* (Universidade Federal de São Carlos, 2014).
56. Barreto, N. R. *Modelo de mecanismo de ruptura da vertente pelas descontinuidades hidráulicas em latossolo no sítio urbano de Santa Teresa, Espírito Santo* (Universidade Federal do Espírito Santo, 2016).
57. Viçosa, F. D. V. C. & Fernandes, M. T. *Avaliação das técnicas de laboratório e campo para realização de ensaios de permeabilidade de solos* (Universidade Federal de Viçosa, 2017).
58. Sisto, F. P. *Estudo da Vulnerabilidade dos Solos na Bacia Hidrográfica do Ribeirão Claro-SP* (Universidade Estadual Paulista, 2016).
59. Reatto, A., Bruand, A., Silva, E. D., Martins, E. S. & Brossard, M. Hydraulic properties of the diagnostic horizon of Latosols of a regional toposequence across the Brazilian Central Plateau. *Geoderma* **139**(1–2), 51–59. https://doi.org/10.1016/j.geoderma.2007.01.003 (2007).
60. Reis, R. P. A., Oliveira, L. H. & Sales, M. M. Sistemas de drenagem na fonte por poços de infiltração de águas pluviais. *Ambiente construído.* **8**(2), 99–117 (2008).
61. Azevedo, G. F., Carvajal, H. E. M. & Souza, N. M. Análise de ameaça de deslizamentos pelo uso de abordagem probabilística aplicada a um modelo de estabilidade de taludes tridimensional. *Geociências (São Paulo).* **37**(3), 655–668 (2019).
62. Elmashad, M. E. & Ata, A. A. Effect of seawater on consistency, infiltration rate and swelling characteristics of montmorillonite clay. *HBRC J.* **12**(2), 175–180. https://doi.org/10.1016/j.hbrcj.2014.12.004 (2016).
63. Gitirana, G. F. N. & Fredlund, D. Statistical assessment of hydraulic properties of unsaturated soils. *Soils Rocks.* **39**(1), 81–95. https://doi.org/10.28927/SR.391081 (2016).
64. Mendes, T. A. *Modelagem física e numérica da infiltração e escoamento de superfícies não saturadas e com cobertura vegetativa* (Universidade de Brasília, 2019).
65. Vaz, E. F. *On the Equilibrium of Suction and Pressure Plate Tests* (Universidade Federal de Goiás, 2019).
66. Araújo, A. G. *Uso de microtomografia e porosimetria para análise da estrutura bimodal de um latossolo vermelho reconstituído* (Universidade Federal de Goiás, 2019).
67. Gribb, M. M., Kodesova, R. & Ordway, S. E. Comparison of soil hydraulic property measurement methods. *J. Geotech. Geoenviron. Eng.* **130**(10), 1084–1095. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:10(1084) (2004).
68. Elfaki, A. A. E. *Correlation between Laboratory and Field Permeability of Soils and Rocks* (University of Khartoum, 2006).
69. Elhakim, A. F. Estimation of soil permeability. *Alex. Eng. J.* **55**(3), 2631–2638. https://doi.org/10.1016/j.aej.2016.07.034 (2016).

Acknowledgements

The authors would like to thank the Instituto Federal de Educação, Ciência e Tecnologia de Goiás (IFG, Federal Institute of Education, Science and Technology of Goias), Universidade de Brasília (UnB, University of Brasilia), Universidade Federal de Goiás (UFG, Federal University of Goias), Université du Québec, Abitibi-Témiscamingue, Canada, North Caroline State University and the coordinator of this research professor Ph.D. Thiago Augusto Mendes.

Author contributions

Conceptualization: W.A.R.S., S.A.S.P., R.F.C., T.A.M.; Methodology: W.A.R.S., S.A.S.P., T.A.M., G.F.N.G., J.F.R.R.; Data curation: W.A.R.S., S.A.S.P., T.A.M., R.F.C.; Writing: W.A.R.S., S.A.S.P., R.F.C., T.A.M.; Supervision: T.A.M., G.F.N.G., J.F.R.R.; Analysis and interpretation of results: W.A.R.S., S.A.S.P., T.A.M., G.F.N.G., J.F.R.R.

Funding

This research did not receive any specific funding.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.