



# Artificial neural networks for predicting soil water retention data of various Brazilian soils

Lucas Broseghini Totola<sup>1</sup> · Kátia Vanessa Bicalho<sup>1</sup> · Wilian Hiroshi Hisatugu<sup>2</sup>

Received: 4 July 2023 / Accepted: 28 September 2023 / Published online: 4 October 2023  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

Knowledge of the soil water retention (SWR) data is necessary for modeling soil water movement and assessing soil water holding capacity and availability. Since direct measurement is often time-consuming and costly, pedotransfer functions (PTFs) have been widely used to predict SWR data from basic soil physical properties. Considering the limited availability of PTFs derived from tropical soils, this paper developed artificial neural networks based on the pseudo-continuous approach (NN-PTFs) to predict SWR data for Brazilian soils. Natural logarithm of soil suction,  $\ln(h)$ , is considered as an extra input parameter in this approach. It enables to predict SWR data at any desired soil suction as it results in more extensive and useful database. The analysis was conducted on a previously compiled hydrophysical database for Brazilian soils representing a variety of soil compositions. The results demonstrated high accuracy and reliability in estimating SWR data, with an overall error of  $0.045 \text{ cm}^3 \cdot \text{cm}^{-3}$ , when incorporating both soil texture (i.e., clay, silt, and sand fractions) and soil structure-related properties (i.e., soil density, particle density and organic matter content) as input parameters. Moreover, the proposed NN-PTFs outperformed PTFs developed for temperate climates, as well as equation-based PTFs derived for specific tropical locals, particularly for weathered soils. The results highlight not only the potential of using NN-PTFs to predict pseudo-continuous SWR curve in preliminary studies, but also their flexibility and the benefits of not limiting the SWR data to a pre-defined function.

**Keywords** Soil-water retention · Artificial neural networks · Unsaturated soils · Brazilian soils

## Introduction

The term soil water retention (SWR) or soil water characteristic (SWC) curve refers to the relationship between the amount of water in the soil and the corresponding soil

suction in the pore water. SWR data are required for modeling the hydraulic and mechanical behavior of unsaturated soils (Karube and Kaway 2001, Fredlund 2006, Angelaki et al. 2023) and understanding the influence of changing environmental conditions at the highly dynamic soil-atmosphere interface of geo-structures (Cui 2022). Knowledge of the SWR data is necessary for modeling soil water movement and assessing soil water holding capacity and availability in soil science, as well as geotechnical and environmental engineering. However, SWR data measurements for wide ranges of soil suction, either in situ or in a laboratory, are often time-consuming and costly (Masrouri et al. 2008; Fredlund and Fredlund 2020) and implicate intensive field sampling campaigns (Montzka et al. 2017). Hence, considerable efforts have been invested in developing indirect methods for estimating SWR data when limited or no test results are available (Botula et al. 2014).

Initially proposed by Bouma (1989), the development of pedotransfer functions (PTFs) has gained prominence since then. These functions provide an empirical relationship

---

Communicated by H. Babaie.

✉ Lucas Broseghini Totola  
lucasbroseghini@gmail.com

Kátia Vanessa Bicalho  
katia.bicalho@ufes.br

Wilian Hiroshi Hisatugu  
wilian.hisatugu@ufes.br

<sup>1</sup> Civil Engineering Department, Federal University of Espírito Santo, Av. Fernando Ferrari, 514, Espírito Santo Vitoria 29075-910, Brazil

<sup>2</sup> Industrial Technology Department, Federal University of Espírito Santo, Av. Fernando Ferrari, 514, Vitoria, Espírito Santo 29075-910, Brazil

between soil hydraulic properties and basic soil properties (Wösten et al. 2001; Chin et al. 2010; Zhang and Schaap 2019). SWR data exhibit a non-linear nature, hysteresis, and are influenced by several factors, including soil type, soil structure and density, and the initial state of the soil sample (Masrouri et al. 2008). Soil physical properties such as the particle size distribution, bulk density (BD) and organic matter content (OM) are often used to predict SWR data (Rawls et al. 1991; Al Majou et al. 2018).

Most methods used for creating PTFs are empirical, with their parameters calibrated using existing hydraulic soil data (Van Looy et al. 2017). Vereecken et al. (2010) highlighted two main categories of PTFs: linear and non-linear regressions, which are popular due to their simplicity; and data mining and exploration techniques, including artificial neural-networks (NNs), k-nearest neighbors (k-NNs), support vector machines, and decision trees. There has been a recent surge in interest in the latter group, as several studies have proved its ability to model the complex soil-water interactions with superior and reliable prediction performance (Botula et al. 2013; Nguyen et al. 2017).

NNs work similarly to the human brain and nervous system. They consist of a series of processing elements called neurons, organized in layers between the input and output layers. The connections between neurons are represented by connection weights, which are fitted using specific learning algorithms. The technique is remarkable for its ability to learn and generalize knowledge (Shahin 2013). NNs are also data-driven, allowing them to uncover implicit relationships between input and output data. This makes them particularly well-suited for handling large databases, a challenge for traditional regression methods (Schaap et al. 2001). Finally, the need for a pre-determined SWR function (i.e., van Genuchten 190) or pre-selected called critical SWR points (Ren et al. 2020) is eliminated.

Several studies can be found in the literature that use NN-PTFs for predicting SWR data (Pachepsky et al. 1996, Schaap and Boultou 1996; Minasny et al. 1999; Schaap et al. 2001; Borgesen and Schaap 2005; Haghverdi et al. 2012, Haghverdi et al. 2014, 2015, 2018; Nguyen et al. 2017; Saha et al. 2018; Pham et al. 2019). There are three commonly used approaches for developing NN-PTFs: point, parametric, and pseudo-continuous. The point approach estimates SWR data at pre-defined suctions values, but it does not provide a continuous SWR curve. The parametric approach estimates model parameters of a pre-defined SWR function, which can express the SWR curve in a closed and continuous form, such as the van Genuchten (1980) model. Some researchers (Tomasella et al. 2003; Vereecken et al. 2010) have pointed out problems in correlating the parameters of these models with soil physical properties and consider

parametric estimations generally less accurate than those from PTFs based on the point approach.

Haghverdi et al. (2012) proposed a called pseudo-continuous approach, which considers the natural logarithmic of suction as an additional input parameter. This addition allows the user to predict water content ( $\theta_h$ ) at the soil suction ( $h$ ) used as a predictor. Therefore, when using a wide range of suctions as inputs, a corresponding range of water content values will be predicted, resulting in a pseudo-continuous SWR curve (Haghverdi et al. 2012). The researchers demonstrate that their approach outperformed parametric NN-PTFs approach and performed slightly better than the point NN-PTFs approach for soils from Iran and Australia. The called pseudo-continuous approach also yielded good results for soils from Turkey and Belgium (Haghverdi et al. 2014, 2018), the United States (Haghverdi et al. 2015), Vietnam (Nguyen et al. 2017), and North America and Europe (Pham et al. 2019).

The NN-PTFs performance is not only influenced by the good correlation between the desired output and the available inputs but is also affected by the quality and extent of the database used in the model's creation and validation (Schaap and Leij 1998; Nguyen et al. 2017). In most tropical soil regions, there is insufficient soil hydraulic data available for PTF development. Therefore, for these areas, it is common to apply PTFs derived from temperate soils, where extensive databases are available (Haghverdi et al. 2012). However, several studies have highlighted limited performance in these cases, mainly due to differences in physical and chemical properties between temperate and tropical soils (Tomasella et al. 2000; Hodnett and Tomasella 2002; Botula et al. 2012; Ottoni et al. 2018, 2019). Residual highly weathered soils from tropical and subtropical regions typically exhibit well-defined macrostructures (inter-aggregate) and microstructures (intra-aggregate) caused by strong weathering and leaching processes. Consequently, these soils may present a bimodal pore-size distribution and SWR curve rather than the conventional unimodal shape (Feuerharmel et al. 2006; Miguel and Bonder 2012). Haghverdi et al. (2020) do not recommend developing parametric PTFs using bimodal variants due to their poor performance, which is associated with their high number of free parameters. Additionally, the clay fraction of temperate soils only minimally covers the variation in the higher clay content typically found in tropical soils (Tomasella and Hodnett 2004).

Barros and Jong van Lier (2014) conducted an extensive review of PTFs for predicting SWR data of Brazilian soils. Most of these studies focused on parametric PTFs that estimates van Genuchten (1980) equation parameters using linear and non-linear regressions. However, limited databases or specific geographical regions have been investigated

(e.g., Reichert et al. 2009; Medrado and Lima 2014), and/or some predictors are not easily accessible (e.g., Hodnett and Tomasella 2002; Tomasella et al. 2003), restricting their applicability.

The advent of HYBRAS (*Hydrophysical Database for Brazilian Soils*) (Ottoni et al. 2018, 2019) has provided a large and consistent hydrophysical data for soils across Brazilian territory. In order to overcome the lack of PTFs development to estimate SWR data representing a variety of soil compositions under tropical and subtropical conditions, this paper aims to: (1) develop NN-PTFs based on the pseudo-continuous approach to estimate SWR data for Brazilian soils using available soil physical properties (e.g., soil texture, bulk density, particle density, total porosity, organic matter content); (2) explore different combinations of input parameters and architectures to improve the accuracy of the proposed models; (3) investigate the NN-PTFs ability to predict a wide range of suction values on the SWR data for various soil textural classes and pedogenic origins; and (4) assess the reliability of the models proposed using an independent validation set and compare their accuracy with the results obtained from PTFs models in the literature, developed for different tropical and temperate regions.

## Materials and methods

### Data collection and preparation for training and prediction

Due to the wide diversity of pedogenesis and soil formation factors, the Brazilian territory exhibits a variety of soils across its continental extension. The HYBRAS (Ottoni et al. 2018; 2019) compiled information on the physical and hydraulic properties of numerous Brazilian soils, sourced from over 30 scientific studies. For this study, an initial compilation of 770 soil samples was undertaken. The selection criteria considered the joint availability of texture information (i.e., mass percentages of sand, silt and clay fractions), along with bulk density (BD), particle density (PD), total porosity ( $n$ ) and organic matter content (OM), all used as input parameters; and soil water content data ( $\theta$ ) measured across a broad suction ( $h$ ) range (0–1500 kPa), used as the output parameter. Given the influence of outliers on the performance of NN-PTFs (Pham et al. 2019), data quality was upheld by applying Tukey's rule to remove outliers from each soil input parameter. Outliers were observed in 33 soil samples characterized by high organic matter content ( $> 8.21\%$ ) and particle density values exceeding 3.08, which were underrepresented within the data set. Moreover, it was also ensured no increase in soil water content values during the drying SWR path. Finally, the assembled data set

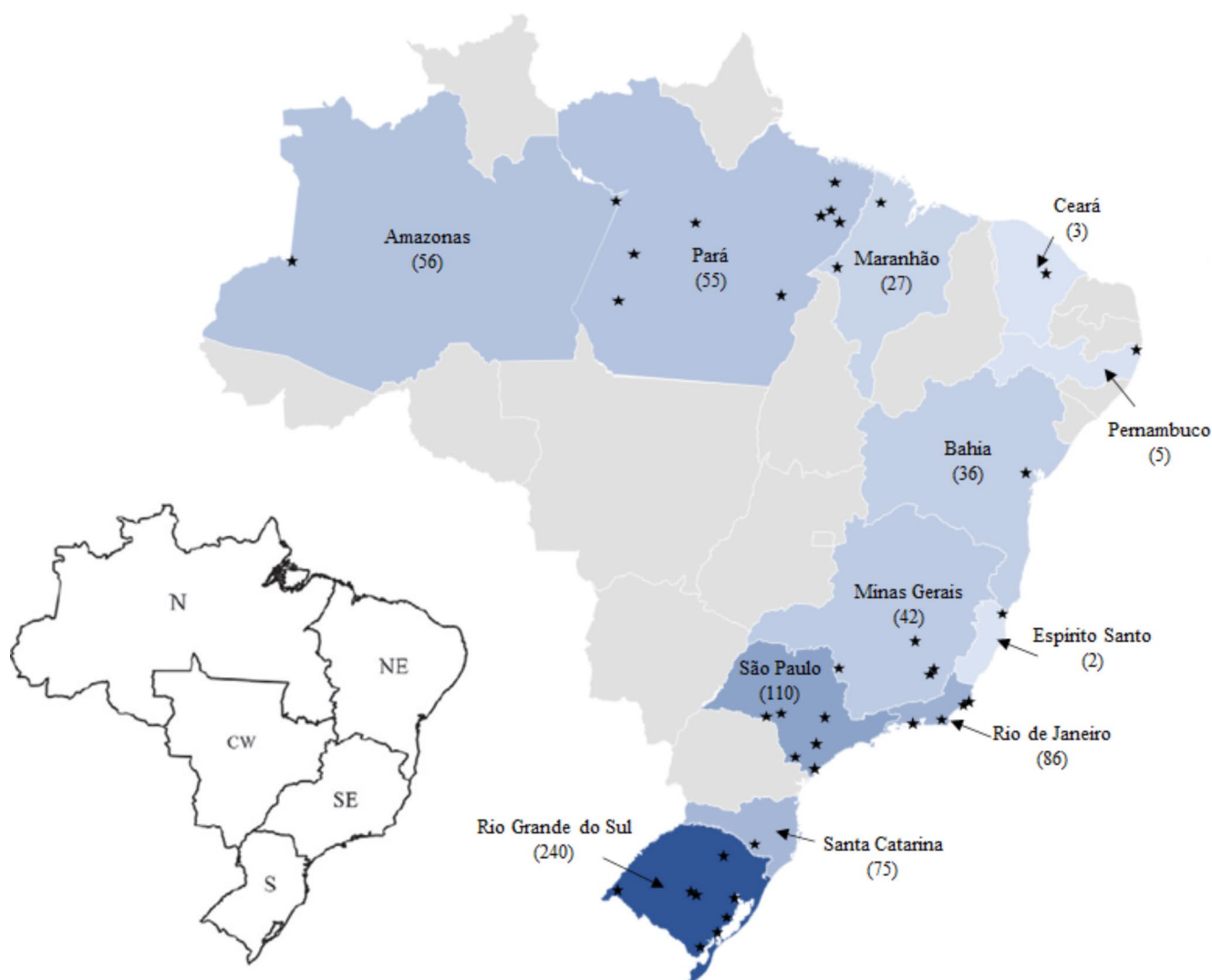
consisted of 737 soil samples, with a total of 5818 experimental points ( $\theta, h$ ).

The input parameters selection for PTFs development is typically based on prior knowledge. As previously mentioned, the SWR behavior is influenced by several factors, and the chosen input parameters are designed to address these influences: soil texture is represented through the distribution ratio of clay, silt and sand contents (SSC); soil structure and initial state of the soil sample are captured by variables such as  $BD$ ,  $PD$ ,  $n$  and  $OM$ ; while the soil stress state is represented by the soil suction natural logarithm  $\ln(h)$ .

The experimental pairs of SWR data ( $\theta, h$ ) were predominantly obtained through methods that directly measured suction and water content (e.g., the pressure plate method). Tests were conducted on undisturbed and initially saturated soil samples (Ottoni et al. 2018, 2019), representing the drying (desorption) SWR data of the soils under investigation. Substantial understanding of unsaturated soil behavior can be accrued by focusing on the drying path as its estimations prove suitable for most geotechnical engineering applications (Fredlund and Fredlund 2020). Soil specimens with volume size within 46 and 102 cm<sup>3</sup> were often used to determine the experimental points ( $\sim 92\%$  of  $\theta$  measurements with size information). Particle size distributions were all within the limits of the American classification system USDA-FAO (clay  $< 0.002$  mm, silt 0.002–0.05 mm, and sand 0.05–2 mm). Bulk density (BD) was determined by the volumetric ring method, while particle density (PD) was primarily determined using the volumetric flask and pycnometer methods. Total porosity ( $n$ ) was calculated as the ratio ( $n = 1 - BD/PD$ ) and organic matter content (OM) was obtained from Walkley and Black (1934) titration method.

The 737 soil samples covers 40 cities across 12 of Brazil's 26 states. Most of the data originates from southern Brazil ( $\sim 43\%$ ), with additional data from Northeast, Southeast and Brazilian Amazonia regions (Fig. 1). The main soil groups within the data set according to the FAO soil classification system (IUSS Working Group WRB 2015), include Planosols (169 samples), Ferralsols (169 samples) and Acrisols (132 samples). The latter two are representative of weathered tropical environments, which predominate in Brazil and account for 60% of the nation's territory (Ottoni et al. 2014). The remaining soils also included Nitisols (97), Gleysols (72), Cambisols (62), Podzols (14), Phaeozems (11), Regosols (6) and Histosols (5). The soil saturated hydraulic conductivity ( $K_s$ ) across these soils ranged from  $1.5 \times 10^{-7}$  to  $3.5 \times 10^{-4}$  m/s.

The available database was split into two distinct sets as follows: (a) the training data set comprised 553 soil samples (75%) and was employed as the basis for calibrating the parameters of the proposed NN-PTFs; and (b) a validation



**Fig. 1** Distribution of the investigated soil samples within Brazilian territory. The darker blue areas correspond to states with higher representativeness within the data set. The map also shows the main

geographical regions of Brazil, namely N-north, NE-northeast, CW-central-western, SE-southeast, and S-south

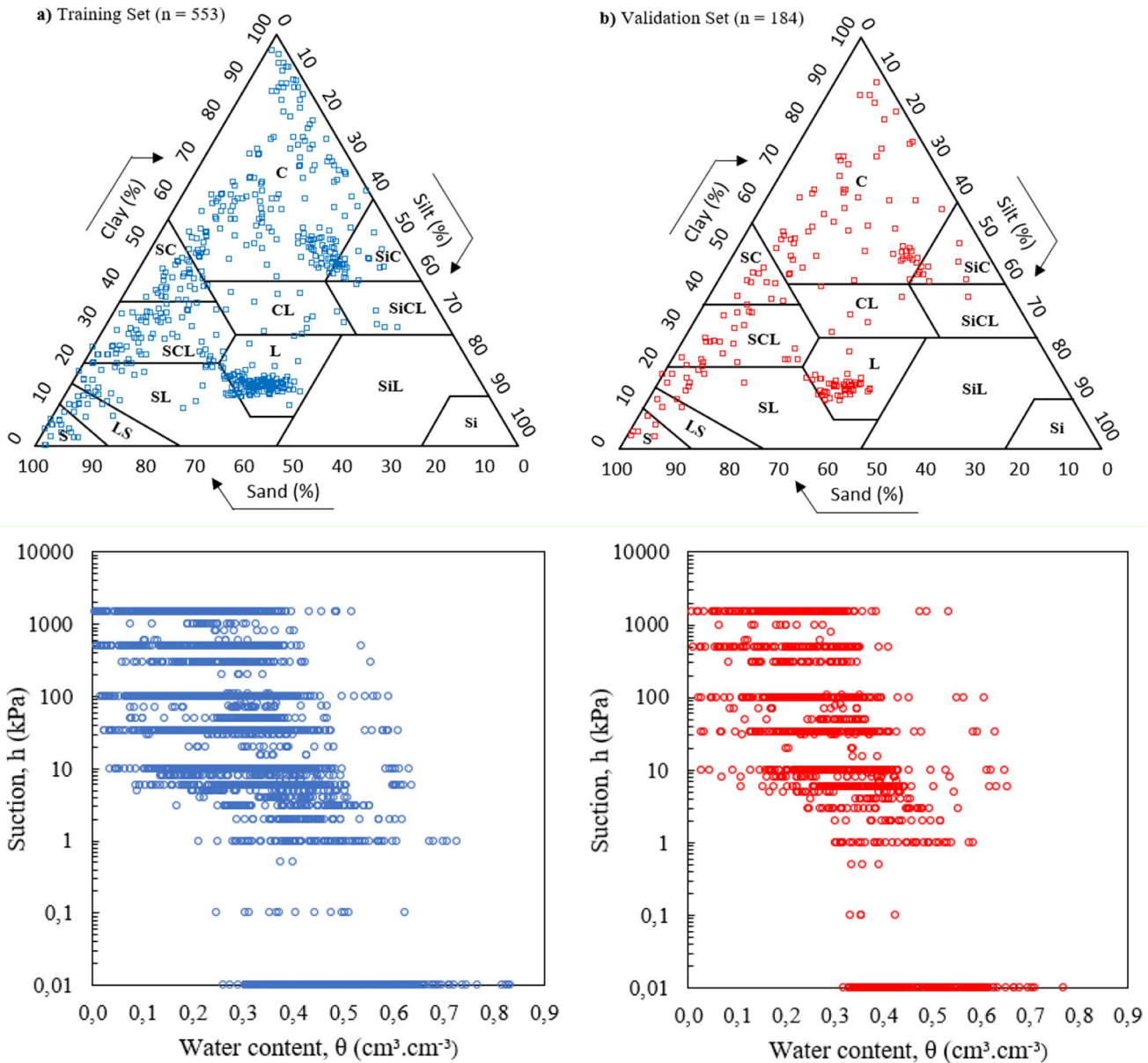
data set composed by the remaining 184 soil samples (25%). This set served the purpose of testing the robustness of the proposed NN-PTFs (i.e., can yield predictions with enough accuracy when exposed to new patterns not encountered during the model's calibration phase). Additionally, the performance of these proposed NN-PTFs for the validation data set was compared with predictions generated by PTFs previously published in the literature. The data division followed a trial-and-error process to guarantee that the calibration and validation sets were statistically homogeneous and thus represent the same statistical population, preventing biased results from the models. And so, t- and F-tests were carried out to examine the null hypothesis of no significant differences in the means and standard deviations ( $\delta$ ) between data sets, respectively. The tests were carried out with a confidence level of 95%. Table 1 describes the minimum, mean,

maximum and standard deviation ( $\delta$ ) values for each soil property across the entire investigated database within both calibration and validating sets.

The soils predominantly occupy the central and left portions of the USDA soil textural triangle (Fig. 2). Across the data set, 10 textural classes are represented, encompassing diverse soil types and hydraulic properties. Notably, a lack of soils with high silt content is observed, a characteristic representation of weathered soils (Otoni et al. 2019). The preponderance of very coarse and very fine particles, relative to particles of intermediate sizes, underscores the critical role that particle packing patterns and soil structure play in shaping tropical soil hydraulic properties (Tomasella et al. 2003). Figure 2 also shows measured SWR points for all soils within each set, once again highlighting the variability of soils and curves across the investigated data.

**Table 1** Statistical description of soil properties for the whole training and validation data sets

Soil Properties	Training Set (n = 553)				Validation Set (n = 184)			
	Minimum	Mean	Maximum	$\delta$	Minimum	Mean	Maximum	$\delta$
Clay (%)	0.00	35.01	96.00	22.19	3.00	34.15	89.00	20.87
Silt (%)	0.00	23.82	60.84	15.07	0.00	23.32	53.90	14.75
Sand (%)	0.40	41.17	97.99	23.64	1.60	42.53	95.73	22.90
Bulk Density ( $\text{g.cm}^{-3}$ )	0.45	1.39	2.01	0.27	0.56	1.41	1.82	0.25
Particle Density ( $\text{g.cm}^{-3}$ )	2.18	2.58	3.08	0.13	2.18	2.57	3.08	0.14
Organic Matter (%)	0.08	2.27	8.21	1.60	0.17	2.26	7.87	1.76
Total Porosity (n)	0.24	0.46	0.83	0.11	0.26	0.45	0.75	0.10



**Fig. 2** Distribution of soil samples regarding clay, silt, and sand percentage within the USDA textural triangle for (a) training set and (b) validation set. The soil texture classes are clay (C), sandy clay (SC), clay loam (CL), sandy clay loam (SCL), sandy loam (SL), loamy sand

(LS), sand (S), loam (L), silty clay (SiC), silty clay loam (SiCL), silt loam (SiL) and silt (Si). SWR measured data of all samples are also shown for each set

**Table 2** Pearson's correlation coefficient between input soil properties for the investigated database

Input Parameters	%Clay	%Silt	%Sand	Bulk Density	Particle Density	Organic Matter	Total Porosity
%Clay	1						
%Silt	-0.24	1					
%Sand	-0.79	-0.41	1				
Bulk Density	-0.70	0.04	0.63	1			
Particle Density	0.14	-0.22	0.01	-0.09	1		
Organic Matter	0.13	0.32	-0.32	-0.45	-0.06	1	
Total Porosity	0.70	-0.10	-0.59	-0.97	0.31	0.43	1
ln(h)	0.03	0.01	-0.04	-0.01	0.01	0.03	0.02

The high correlation among input parameters could also be a reason to the inferior performance of NN models (Pham et al. 2019). Pearson correlation coefficient ( $r$ ) was used to identify the correlations among the investigated parameters within the data set, as shown in Table 2. From the obtained results, it's noteworthy that only one pair (BD – n) showed a high correlation value ( $r = -0.97$ ). The concomitantly use of these properties was subsequently evaluated. The other pairs had relatively lower  $r$  values, thus indicating these properties were suitable for estimating SWR points.

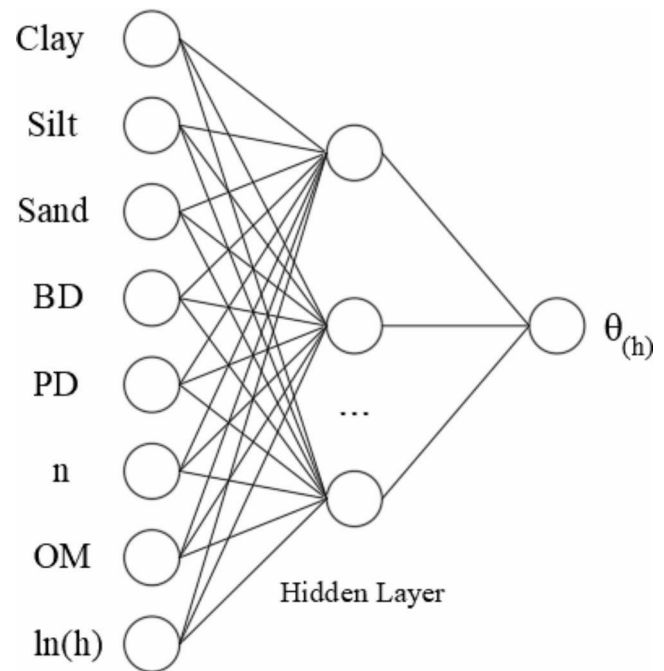
## Development and evaluation of NN-PTFs for SWR data prediction

### NN-PTFs development

The development of NN-PTFs includes several steps, including selecting appropriate model input and output soil parameters, determining model structure and architecture, calibrating model parameters using a learning algorithm, optimizing the model, and validating its performance on data not used during the training phase. The adopted methodology is discussed hereafter. A further description of NNs operation is beyond the scope of this paper.

For this study, the multilayer perceptron (MLP) feed-forward type was chosen and implemented within the *Matlab* environment through the *nftool* and *nntool* interfaces. In its simplest form, the MLP comprises at least three layers: an input layer, an output layer, and one or more hidden layers in between. The study utilized only one hidden layer as analyses revealed that incorporating more than one hidden layer did not impact NNs performance for the investigated data set. The tangent hyperbolic and linear activation functions were respectively assigned to the hidden and the output layer. By employing a nonlinear activation function in the hidden layer, the NN gain the ability to address complex and non-linear problems (Maier et al. 2010), as seen within the SWR data.

The network geometry choice is a challenge for NN-PTFs implementation due to lack of an exact approach. The optimal model structure needs to be determined from trial and error on an iterative process (Maier et al. 2010). Using



**Fig. 3** Topology of the proposed neural-network based pedotransfer function (NN-PTF). BD: bulk density; PD: particle density; n: total porosity; OM: organic matter content; ln(h) where  $h$  is the soil suction (kPa);  $\theta(h)$ : soil water content at  $h$

too few neurons in the hidden layer leads to an excessively simple model that fails to capture all patterns within the data set (underfitting). Conversely, when too many neurons are used, the NN-PTF model learns training set particularities, leading to the memorization of input-output relationships instead of generalization (overfitting) (Shahin 2013). Both underfitting and overfitting results in poor predictions for new data sets. Hence, the number of neurons in the hidden layer ranged from 4 to 14 for the NN-PTFs developed in this study.

The pseudo-continuous topology, as illustrated in Fig. 3, was adopted for this study. This model is similar to the approach introduced by Haghverdi et al. (2012) and consists of adding the natural logarithm of soil suction  $ln(h)$  as an input parameter. Only one output neuron is required, corresponding to the measured water retention point at the given suction. The combination of samples with different measured

water retention points is an advantage over the point PTFs, as it results in more extensive and useful database.

A hierarchical model is then proposed based on the stepwise method (Maier and Dandy 2000). This approach involves estimating water retention points using either limited or more detailed predictors sets. The first model proposed (M1) is texture-based and employs the content of clay, silt and sand (SSC) as inputs, along with  $\ln(h)$ . From this model, the other input parameters (i.e., bulk density, organic matter content, particle density and total porosity) were included one by one, as shown in Table 3. The input parameters were normalized within the activation function range [-1.1] so that the different scales within the input parameters do not affect the model's performance (Pham et al. 2019).

### Models optimization

Models' optimization refers to the use of learning algorithms to minimize an error function by updating NN parameters, such as the weights and biases. In this study, it is carried out using the Levenberg-Marquardt (LM) training algorithm (Hagan and Menhaj 1994). The LM algorithm uses both descent gradient and Gauss-Newton techniques, combining the stability of the first and the convergency speed of the latter – which is achieved by removing the second-order derivatives of the total error function calculation (hessian matrix) and using the jacobian matrix ( $J$ ) instead (Yu and Wilamowski 2011). In an iterative process, the vector containing NN parameters  $w$  is updated after each epoch as follows:

$$w_{k+1} = w_k - [J_k^T J_k + \mu_k I]^{-1} J_k^T e_k \quad (1)$$

where  $k$  is the epoch number,  $J$  is the Jacobian matrix containing first derivatives of the NN errors with respect to the weights and biases,  $\mu$  works as a learning rate and  $e$  is a vector of NN errors.

The learning rate  $\mu$  guides the search for the global minimum on the error surface. In this study, the error surface is defined as the mean squared error (MSE) calculated from estimated results and target outputs as follows:

$$MSE = \frac{1}{N} \sum_{i=1}^N [\theta_{i,est} - \theta_{i,obs}]^2 \quad (2)$$

where  $N$  is the number of data inserted in the model;  $\theta_{i,est}$  and  $\theta_{i,obs}$  denotes the estimated and measured (or target) water retention values, respectively.

The early-stopping technique is also employed as it stops the training process whenever errors on the testing

set increases, thus preventing overfitting. Hence, the 553 soil samples from the training set were randomly divided into two subsets: 80% for training and 20% for testing. The training set is used to adjust NN parameters. Throughout the training phase, errors on the testing set are continuously monitored at each training iteration. The training process is halted as soon as the error for that set increases. Upon completion of the training, the set of NN parameters associated with the better performance was stored and applied to the validation set (184 soil samples). The validation set serves as an independent data set not used in the model building process. Its purpose is to evaluate the model's capacity to generalize within the limits defined by the training data in a robust manner (Shahin 2013). To quantify uncertainties, the training process was repeated a minimum of eight times across the various scenarios under investigation. Default values included a maximum number of iterations set to 1000, an initial learning rate ( $\mu$ ) of 0.001 and maximum validation epoch failures in a row equal to 6.

### Selection of previous published PTFs

Previously published PTFs were selected and evaluated for predicting SWR data of HYBRAS soils for the validation set. The selection criteria considered PTFs that presented characteristics similar to those developed in this study, including input soil properties used, the extent of the data set used for calibration, and the applicability index, defined as the ratio between the number of samples to which a PTF could be applied and the total number of samples in the data set. The four chosen PTFs are as follows: the full model proposed by Medrado and Lima (2014), developed for Brazilian savanna tropical soils; the model proposed by Reichert et al. (2009) for soils from the South Region of Brazil; and two models proposed by Schaap et al. (2001), derived from temperate soils from North America and Europe. Medrado and Lima (2014) proposed a non-linear equation based PTF to estimate the four parameters of the van Genuchten (VG) equation (van Genuchten 1980), which is commonly used to describe the SWR function. Schaap et al. (2001) also predicted VG equation parameters but used NN-PTFs to estimate SWR data. The program *Rosetta* was developed by using a hierarchical structure that allows the use of soil samples with both limited and detailed information. In this study, the proposed models H2 and H3 are applied. Reichert et al. (2009) derived a point-PTF using multiple regression to estimate SWR data at predetermined suctions. The four PTFs selected from the literature and their respective database information are described in Table 3, along with the five NN-PTFs developed in this study using NN.

**Table 3** Information of the investigated and proposed NN-PTFs to predict SWR data and their calibration database values

Model Notation	Inputs	Outputs	n*	Geographical Domain	Clay %	Silt %	Sand %	OM	PD (g.cm <sup>-3</sup> )	BD	AP %	
Proposed NN-PTFs:				Brazil	See Table 1							100
M1	SSC, ln(h)	$\theta_h$	537									
M2	SSC, BD, ln(h)											
M3	SSC, BD, OM, ln(h)											
M4	SSC, BD, OM, PD, ln(h)											
M5	SSC, BD, OM, PD, n, ln(h)											
PTFs from Literature:												
RoSSC	SSC	$\theta_r, \theta_s, n_v, \alpha$ from the VG (1980) equation	2134	North America and Europe	0–89	0–82	0–100		NA*	NA*	100	
RoSSCBD	SSC, BD											
Savanna	SSC, BD, OM, PD, n		1092	Brazilian Savanna (cerrado)	4.8–91.5	0–55	1.2–90	0–7.8	NA*	0.6–1.92	99.5	
South-BR	SSC, BD, OM	$\theta_h$ for h=6, 10, 33, 100, 500 and 1500 kPa	725	Rio Grande do Sul (south Brazil)	1–82	1–78	1–99	0–10	1.96–3.22	0.86–1.8	96.7	

\*NA: data not available

## Statistical evaluation criteria

The accuracy of a PTF is defined by the correspondence between observed and estimated values during the training phase, while the reliability is assessed by applying the model to an independent testing set (Wösten et al. 2001). In this study, the root mean squared error (RMSE) was chosen as the most suitable statistical performance metric for evaluating the developed NN-PTFs models. Additionally, the coefficient of determination ( $R^2$ ) was employed to identify low correlations, while the mean error (ME) indicated overestimation ( $ME > 0$ ) or underestimation ( $ME < 0$ ) tendencies. The NN-PTFs proposed in this study underwent evaluation using both the training and validation data sets. The performance of the PTFs selected from the literature was assessed only for the validation set for comparison purposes. Results were deemed satisfactory when RMSE and ME values were close to zero, and  $R^2$  approached 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\theta_{i,obs} - \theta_{i,est})^2}{N}} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\theta_{i,obs} - \theta_{i,est})^2}{\sum_{i=1}^N (\theta_{i,obs} - \bar{\theta}_{obs})^2} \quad (4)$$

$$ME = \frac{1}{N} \sum_{i=1}^N (\theta_{i,est} - \theta_{i,obs}) \quad (5)$$

where  $\bar{\theta}_{obs}$  is the average of the measured water retention data.

## Results and discussions

### Importance of NN geometry and input variables

The evaluation of the NN-PTFs approach adopted in this study, quantified by RMSE and  $R^2$ , is summarized in Table 4. All scenarios exhibited reasonable accuracy in predicting SWR data. RMSE values ranged from 0.037 to 0.056 for the training set, and from 0.045 to 0.057 for the validation set. The relatively high values of  $R^2$  (0.772 to 0.854) indicate that a large extent of the SWR data variability can be explained by the NN-PTF models proposed using available soil physical properties (i.e., soil texture, BD, n, PD and OM content). The obtained results also revealed an effect on the model's performance of both neural network geometry and choice of input parameters. The simplest model based solely on soil texture (M1) exhibited the lowest performance, while predictions improved when soil structure information was incorporated into the input parameters.

The adopted NN-PTFs performance improved from model M1 to model M4. Thus, the inclusion of more input parameters continuously enhanced efficiency, which agrees



**Table 4** Performance of NN-PTFs developed in this study based on the pseudo-continuous approach for the training and validation sets

Model	Number of Neurons		Training Set			Validation Set				
			RMSE		R <sup>2</sup>		RMSE		R <sup>2</sup>	
	Mean	$\delta$	Mean	$\delta$	Mean	$\delta$	Mean	$\delta$	Mean	$\delta$
<b>M1</b>	4	0.056	0.001	0.009	0.787	0.001	0.057	0.001	0.772	0.008
	7	0.053	0.001	0.004	0.812	0.001	0.055	0.001	0.788	0.004
	10	0.052	0.000	0.003	0.821	0.001	0.054	0.001	0.791	0.005
	14	<b>0.050</b>	<b>0.000</b>	<b>0.003</b>	<b>0.832</b>	<b>0.001</b>	<b>0.054</b>	<b>0.001</b>	<b>0.794</b>	<b>0.005</b>
<b>M2</b>	4	0.048	0.001	0.007	0.843	0.002	0.049	0.002	0.827	0.012
	7	0.045	0.001	0.007	0.861	0.002	<b>0.048</b>	<b>0.002</b>	<b>0.839</b>	<b>0.010</b>
	10	0.043	0.000	0.003	0.876	0.001	0.048	0.001	0.837	0.010
	14	<b>0.041</b>	<b>0.001</b>	<b>0.005</b>	<b>0.886</b>	0.001	0.049	0.001	0.831	0.008
<b>M3</b>	4	0.047	0.001	0.008	0.851	0.001	0.049	0.001	0.831	0.008
	7	0.044	0.001	0.006	0.867	0.001	<b>0.047</b>	<b>0.001</b>	<b>0.844</b>	<b>0.009</b>
	10	0.042	0.001	0.004	0.884	0.001	0.047	0.001	0.842	0.007
	14	<b>0.039</b>	<b>0.001</b>	<b>0.005</b>	<b>0.898</b>	0.002	0.049	0.002	0.831	0.012
<b>M4</b>	4	0.046	0.001	0.005	0.856	0.001	0.048	0.001	0.833	0.007
	7	0.042	0.001	0.006	0.882	0.001	<b>0.045</b>	<b>0.001</b>	<b>0.854</b>	<b>0.006</b>
	10	0.039	0.001	0.005	0.896	0.002	0.048	0.002	0.836	0.013
	14	<b>0.037</b>	<b>0.001</b>	<b>0.004</b>	<b>0.909</b>	0.005	0.052	0.005	0.812	0.032
<b>M5</b>	4	0.045	0.001	0.005	0.861	0.001	0.048	0.001	0.838	0.007
	7	0.042	0.000	0.002	0.883	0.002	<b>0.047</b>	<b>0.002</b>	<b>0.842</b>	<b>0.012</b>
	10	0.039	0.001	0.003	0.897	0.003	0.048	0.002	0.836	0.013
	14	<b>0.037</b>	<b>0.001</b>	<b>0.005</b>	<b>0.908</b>	0.004	0.049	0.004	0.830	0.024

with previous publications (Schaap and Leij 1998; Schaap et al. 2001, Minasny and McBratney 2002; Vereecken et al. 2010). The RMSE dropped by 0.015 and 0.008 for training and validation sets, respectively, when BD was added to the initial model. For the incorporation of PD and OM as input variables (models M3 and M4), the decrease rate of RMSE was less significant, i.e., 0.004 and 0.003 for training and validation sets, respectively. So, the best scenario was observed for M4 with 7 hidden neurons (RMSE = 0.045 and  $R^2 = 0.854$ ). As tropical and subtropical soils typically have high clay contents with specific mineralogical nature, and both factors strongly affect the soil water retention capacity, the selection of input parameters should include soil properties representative of both soil texture and structure to define their soil hydraulic behavior.

The performance improvement with the addition of  $n$  was not clear (model M5), possibly due to the known physical relationship between  $n$  and  $BD$ , along with the high correlation between these properties as indicated in Table 2. Therefore, the efficiency gained through increasing input parameters can be affected by the database information, such as the interdependence between these input parameters.

In terms of training set, all models performed better with more hidden neurons. When the number of neurons increased from 4 to 14, the RMSE dropped by 0.006 (M1) up to 0.009 (M4). Nevertheless, this trend was not observed in the validation set, which assesses model's reliability, suggesting overfitting. When hidden neurons were increased from 4 to 7, RMSE decreased for all models, culminating in the optimal performance for models M2, M3, M4, and M5. Using more than 7 hidden neurons was only efficient for M1, though even its best performance remained inferior to the other models. Moreover, the variability within the results increased as the number of neurons increased, as indicated by the standard deviation.

Overall, the findings on the proposed NN-PTFs accuracy (testing set) and reliability (validation set) are comparable to those of previously developed PTFs based on the pseudo-continuous approach, where reported RMSEs range from 0.027 to 0.056  $\text{cm}^3 \cdot \text{cm}^{-3}$  (Haghverdi et al. 2012, Haghverdi et al. 2014, Haghverdi et al. 2015, Haghverdi et al. 2018; Nguyen et al. 2017; Pham et al. 2019). Haghverdi et al. (2015) related this variation to the

heterogeneity within the data sets, with models developed for small local data sets performing better than those derived for data sets with a variability of soil geological origins.

Based on the statistical metrics, models M1 to M4 with better performance for the validation set were selected for further analyses: M1 with 14 hidden neurons and M2, M3, M4 and M5 with 7 hidden neurons.

### Performance of the proposed NN-PTFs at different suction values

The performance of the proposed NN-PTFs at different suction values (i.e., 0, 6, 10, 33, 100 and 1500 kPa) is summarized in Table 5. These points were chosen based on their representability within the validation data set. The M4 model performed relatively well for any suction value, showing consistency across the entire range of investigated SWR data. The lowest RMSE values can be observed at saturation ( $h = 0$  kPa) when soil structure related parameters were incorporated as input parameters (M2–M5). In fact, a similar error variation pattern was observed across models M2–M5: RMSE remained reasonably low in the wet portion of the curve, increased with the intermediate range, and then decreased again in the dry part. This trend aligns with finding from literature for NN-PTFs based on pseudo-continuous approach (Haghverdi et al. 2012, 2015; Nguyen et al. 2017), as well as both point and pseudo-continuous k-NN-PTFs approaches (Botula et al. 2013, 2014; Nguyen et al. 2017), and parametric PTFs approach (Vereecken et al. 2010). These studies associated this behavior not only to the PTF type but also to the database information and combination of input parameters - all of them used combinations of SSC, BD and OM.

A decreasing pattern of RMSE values (0.057–0.044) can be observed in the model based solely on soil texture data (M1), moving from the wet to the dry range of the SWR. In contrast, with the addition of BD, OM, PD and  $n$  as input parameters, errors were low within the wet range (for  $h = 0$  kPa,  $< 0.032$  and for  $h = 6$  kPa,  $< 0.045$ ). These parameters are important properties that account for water retention variation in the SWR data wet range, which is strongly governed by soil structure. In fact, soil structure plays a crucial role in defining the hydraulic behavior in the macropore

**Table 5** RMSE variation for the validation set at different suction values (h) for the proposed NN-PTFs

Model Notation	Suction (kPa) and number of experimental points						
	0	6	10	33	100	500	1500
	184	162	179	172	177	105	182
M1	0.057	0.046	0.060	0.058	0.055	0.040	0.044
M2	0.032	0.045	0.053	0.054	0.053	0.040	0.046
M3	0.031	0.044	0.053	0.054	0.052	0.039	0.045
M4	0.031	0.041	0.050	0.050	0.049	0.041	0.044
M5	0.032	0.041	0.052	0.052	0.050	0.042	0.044

flow region (Mermoud and Xu 2006; Weynants et al. 2009). The decrease rate of RMSE was less significant within the intermediate range (i.e., between  $h = 10$  and 100 kPa), with values ranging from 0.049 to 0.054 for models M2 to M5. For the dry range of the SWR data ( $h > 500$  kPa), which is highly influenced by adsorption forces, the RMSE values found were similar across all proposed models ( $\sim 0.044$ ). Given that the clay fraction predominates within the fines of the soils under investigation, clay mineralogy should be a significant factor when predicting the hydraulic behavior of these soils, and should be incorporated on PTFs development, as pointed out by Tomasella et al. (2000), Hodnett and Tomasella (2002), Botula et al. (2013) and Reichert et al. (2020). Additionally, since water retention is controlled by distinct independent properties across different ranges of the SWR data, different combinations of input parameters could be used to achieve more accurate estimates (Tomasella et al. 2003).

### Performance of the proposed NN-PTFs at different soil textural classes and groups

Figure 4 illustrates the average performance of the proposed NN-PTFs (M1–M4) across different soil textural classes and soil groups according to the FAO system. Additionally, the percentage representation of each soil class/group within the validation set is plotted. Some groups are not represented due to the lack of data. Vereecken et al. (2010) suggested that the percentage share of textural classes within the data set could impact model performance. Both derived from temperate data sets, Haghverdi et al. (2014) and Haghverdi et al. (2018) associated the highest accuracy of their models with textural classes with the highest number of soil samples within the training set. Such behavior is discernible within the investigated Brazilian soil data set when considering the FAO system classification, as the highest errors arise for groups with the lowest amount of data (i.e., unweathered soils groups of Gleysols and Podzols). In contrast, the lowest errors are observed for Planosols ( $< 0.029$ ). The models were also able to capture the behavior of weathered soils (i.e., Acrisols, Ferralsols and Nitisols) relatively well, a challenge for PTFs derived from temperate soils and even for tropical soils (Otoni et al. 2018). In fact, as these groups presented a large amount of data within the HYBRAS, this reflects the capacity of the proposed models to represent weathered soils better than soils from other pedogenic environments.

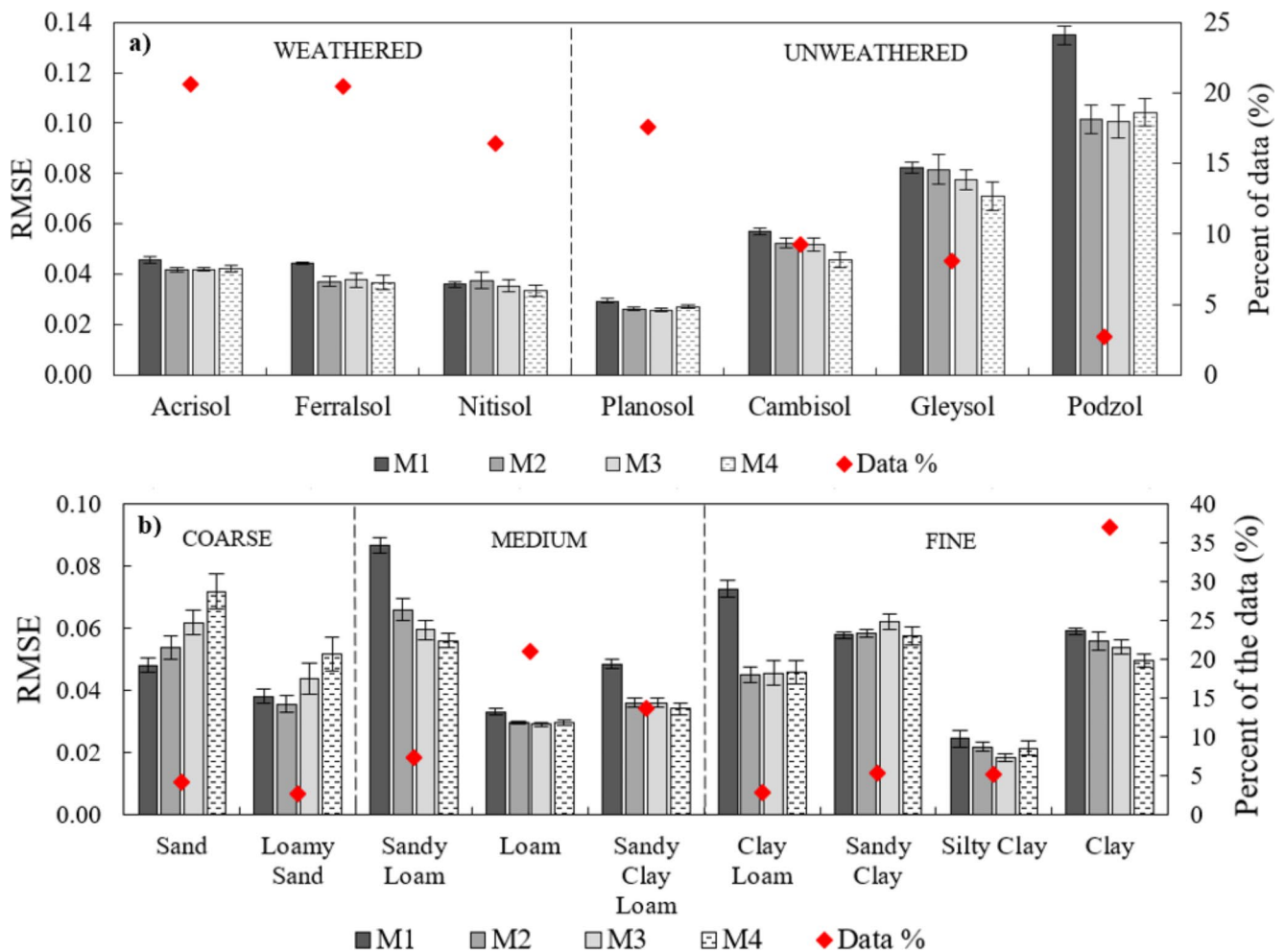
However, this pattern, wherein PTFs accuracy inversely correlates with the number of soil samples in each group, does not hold for the studied data set when contrasting performance across soil textural classes (Fig. 4b). Sandy Loam class accrues the highest error ( $> 0.056$ ), while Silty Clay

class boast the lowest ( $< 0.025$ ). Even with the highest amount of data, the Clay soil texture class exhibits RMSE values higher than 0.050. As a similar trend of errors is observed across models M2–M4, the higher errors within a class may be related to either SWR data heterogeneity for soils within that class or a lower correlation between input parameters and SWR data for the class (Haghverdi et al. 2014). Also, as expected, the model based on soil texture (M1) satisfactory estimated SWR data from HYBRAS coarse soils, while soil structure-related properties were essential for medium and fine soils, as errors decreased for these classes within the NN-PTFs models M2–M4.

Overall, the consideration of soil texture classes itself was insufficient to evaluate the proposed models as the water retention for tropical soils is highly dependent on soil mineralogy, and this characteristic varies among soils and their respective horizons. So, the consideration of soil pedogenic environments should also be considered for database construction and the evaluation of PTFs' performance.

### Comparison of performance of the proposed NN-PTFs models with published PTFs

Upon comparison of the performance of the evaluated PTFs in this study using the complete validation data set (184 samples) (Table 6), it becomes evident that the proposed NN-PTF models (M1–M4) exhibit the lowest estimation errors (RMSE and ME). However, the conclusions here should be taken carefully, considering the published PTFs do not use the same inputs and are derived from different databases. The M1 model proposed requires the same input properties as the RoSSC PTF, while the proposed M2 model is comparable to the RoSSCBd PTF. In both cases, the proposed NN-PTFs produced more reliable SWR data estimates than the use of models derived from temperate soils (RoSSC and RoSSCBd). This assessment holds true when examining the entire validation set, and when distinguishing between weathered and unweathered soils, which was also found by Tomasella et al. (2000), Reichert et al. (2009), Botula et al. (2012) and Otoni et al. (2018). The underestimating trend of both Rosetta models can be somewhat due to the difference in soil properties between HYBRAS data set, with a frequent occurrence of high clay contents which presents peculiar mineralogy from weathered environments, and the data set used to develop the Rosetta PTFs, as observed by Otoni et al. (2018). Utilizing data from different locations for calibration and testing results in weaker predictions compared to those derived from calibration and testing with similar pedogenic origin and soil properties information. In fact, both RoSSC and RoSSCBd models exhibited better estimates for coarse soils, typical within the deriving



**Fig. 4** Relationship between root mean square error (RMSE) and (a) soil groups according to FAO soil classification system and (b) soil textural classes. The diamonds represent the percent of the data (%)

**Table 6** Root mean square error (RMSE) and mean error (ME) for the evaluated pedotransfer functions (PTFs) for the validation set, and RMSE values for weathered and unweathered soils

Parameter	Proposed NN-PTFs				Temperate PTFs		Tropical PTFs	
	M1	M2	M3	M4	RoSSC	RoSSCBD	Savanna	South-BR
Complete Validation Set (144 samples):								
RMSE	0.054	0.048	0.047	0.045	0.072	0.073	0.066	0.070
R <sup>2</sup>	0.79	0.84	0.84	0.85	0.69	0.73	0.72	0.67
ME	-0.13	-0.17%	-0.05%	0.09	-2.32%	-1.66%	0.48	-4.67%
Weathered soils (96 samples):								
RMSE	0.043	0.039	0.039	0.038	0.059	0.069	0.066	0.054
Unweathered soils (88 samples):								
RMSE	0.066	0.057	0.056	0.054	0.087	0.078	0.065	0.088

database of these PTFs, and the errors increased for clayey textural classes.

The proposed models also outperformed both investigated PTFs derived from tropical soils – Savanna and South-BR. Although the errors of these two PTFs were comparable in terms of RMSE (0.066 and 0.070, respectively), an assessment of the mean error (ME) highlighted

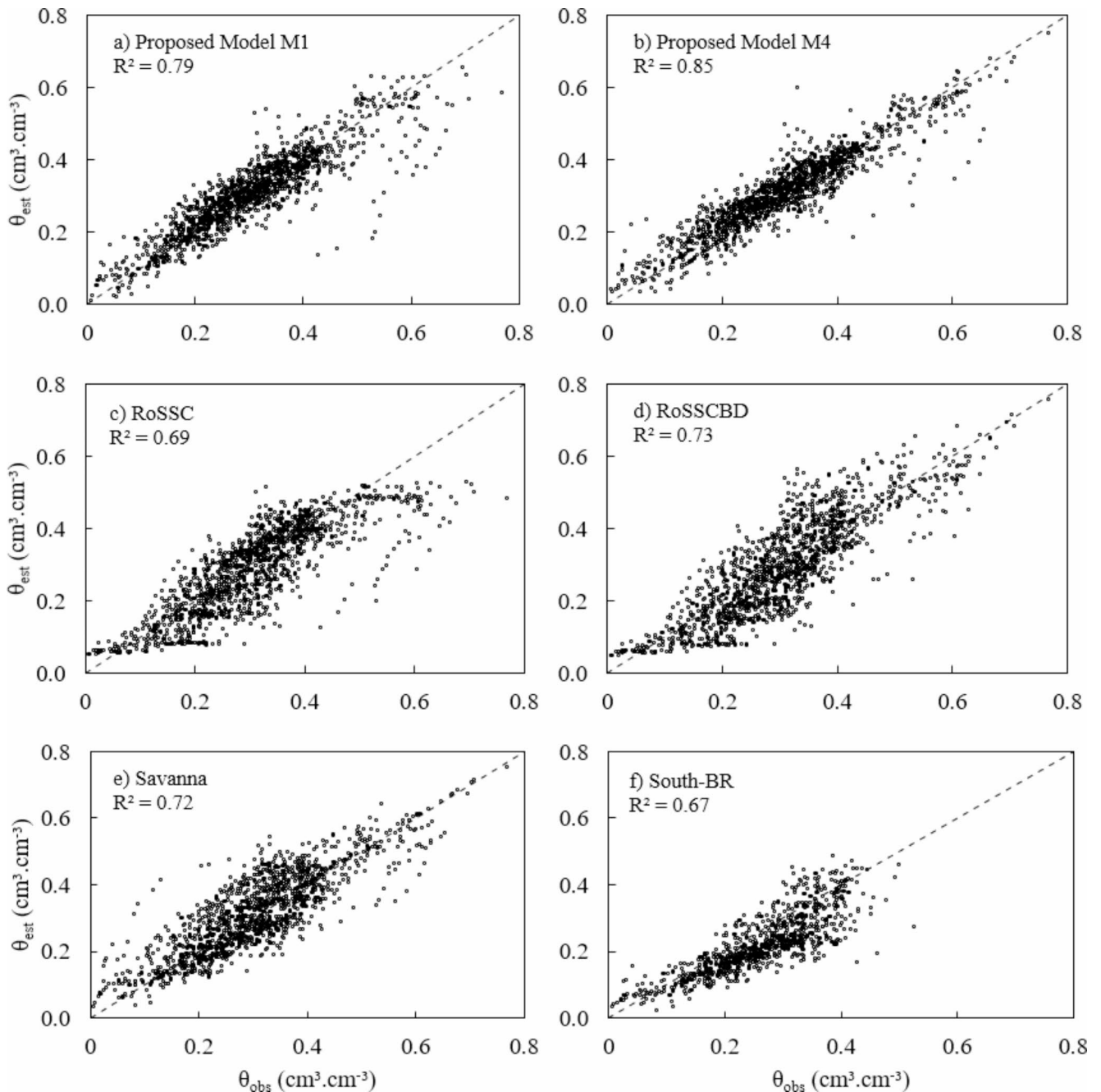
an underestimation trend for the South-BR model (ME = –4.67%). The better performance observed for the proposed NN-PTFs are especially important, considering these tropical PTFs were derived for soils with similar characteristics to those utilized in this study: Savanna was developed for Brazilian savanna soils, predominantly Ferralsols, while the South-BR model used soils from Rio Grande do Sul, which

are majority within the investigated database. Also, Savanna and South-BR PTFs employed similar input parameters to the proposed models M3 and M4, except for  $\ln(h)$ .

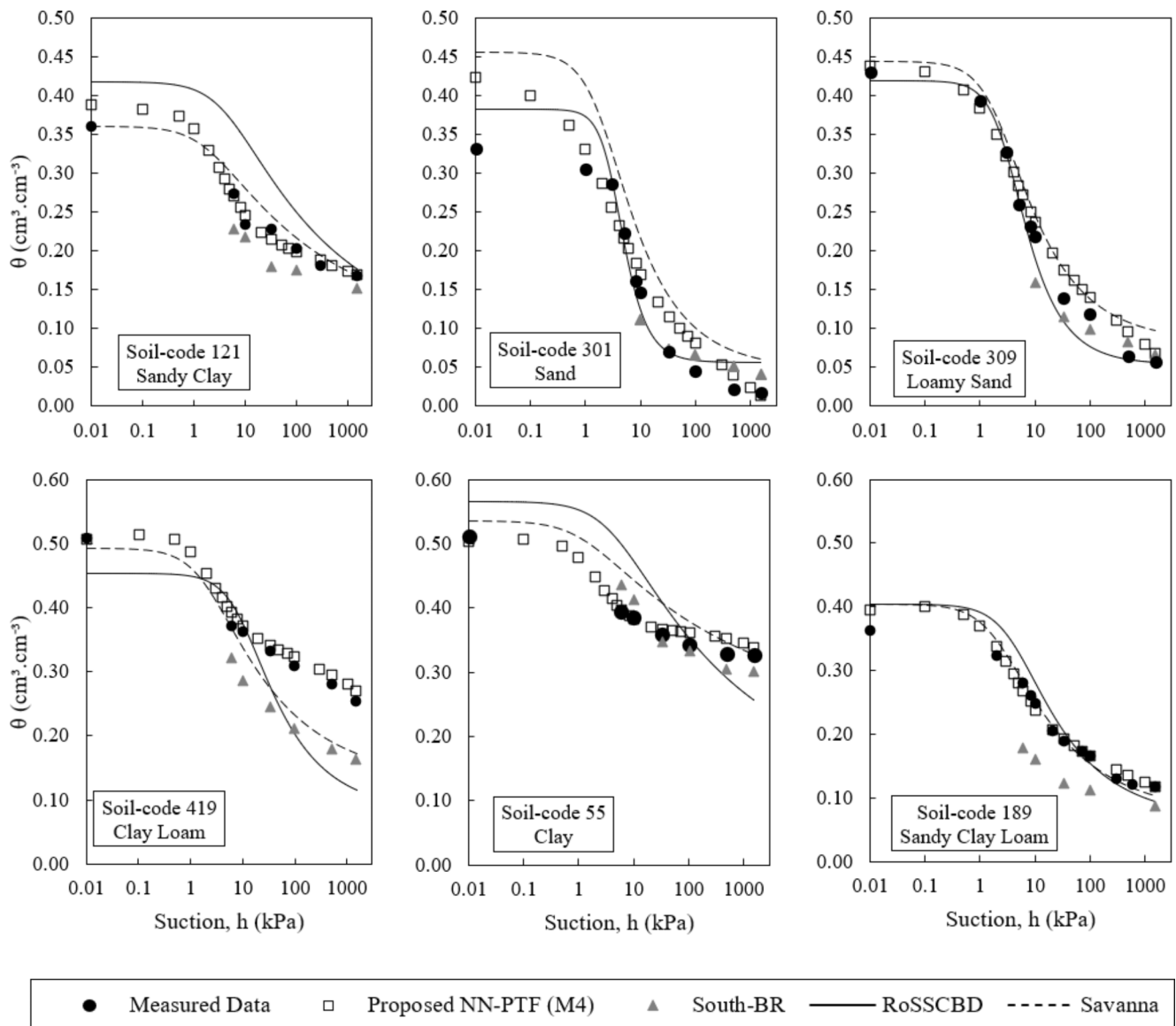
Figure 5 illustrates the correlation between the measured ( $\theta_{obs}$ ) and estimated ( $\theta_{est}$ ) water retention values for the NN-PTF models M1 and M4 proposed in this study, as well as the models RoSSCBd, Savanna and South-BR from the literature. The texture-based models (M1 and RoSSC) underestimated higher  $\theta$  values ( $> 0.50 \text{ cm}^3.\text{cm}^{-3}$ ), as well as the South-BR model. Conversely, models incorporating

soil structure input parameters exhibited agreement between  $\theta_{obs}$  and  $\theta_{est}$  across the entire investigated range. Overall, the errors increased for soils with high organic matter content ( $> 7\%$ ) and low bulk density soils ( $< 0.8 \text{ g.cm}^{-3}$ ).

The SWR data estimated by the proposed NN-PTF model M4 and the corresponding experimental data across different textural classes are illustrated in Fig. 6. The predictions from models RoSSCBd, Savanna and South-BR PTFs are also presented. It is noteworthy to observe the differences between the approaches used to establish the PTFs,



**Fig. 5** Relationship of water retention points measured ( $\theta_{obs}$ ) for different PTFs: **(a)** Proposed NN-PTF M1, **(b)** Proposed NN-PTF M4, **(c)** RoSSC, **(d)** RoSSCBd, **(e)** Savanna and **(f)** South-BR, and experimental data ( $\theta_{est}$ ) for the validation data set



**Fig. 6** Predicted soil water retention curves given by NN-PTFs model M4 developed in this study and PTFs from the literature. The black dots represent the experimental data of each soil sample

which refers to the continuity and shape of the SWR curve. Since the M4 model obtains a (pseudo) continuous and freely shaped curve, enabling predictions of SWR data at any desired suction, it exhibits a good agreement between observed and estimated water retention values. On the other hand, the point-based South-BR PTF predicts water content only at predefined suctions, and the underestimation pattern indicated by the ME analyses is clear. Lastly, the parametric PTFs (RoSSCBD and Savanna) are based on a closed-form expression, which might not be the best fit for all soil samples within a data set.

## Conclusions

NN-PTFs based on the pseudo-continuous approach were developed to estimate SWR data of various Brazilian soils under tropical and subtropical conditions. Since PTFs developed for predicting SWR data for Brazilian soils are still scarce or limited, this paper provides a robust NN-PTF model by using basic and commonly available soil information from a database containing soil samples of diverse pedogenesis.

A hierarchical structure of NN-PTFs is proposed to predict SWR data from different combination of input parameters. In general, better performance is achieved when using

both soil texture and soil structure-related properties as input parameters.

The proposed NN-PTFs outperformed previously published PTFs for temperate soils, which is explained by the differences between soils used to derive and apply these PTFs. Hence, arbitrary use of PTFs may lead to inconsistent results and should be avoided. Previously published PTFs proposed for tropical soils based on point and parametric approaches were also outperformed by the proposed NN-PTFs. The adopted pseudo-continuous approach does not limit a particular fixed function for the SWR curve, as seen in parametric PTFs based on closed-form expressions. It can be used in a hierarchical way, depending on the availability of limited or more detailed data, and it facilitates combining data where water content have been determined at different suctions, resulting in a more extensive dataset compared to the point approach. Finally, it shows flexibility by allowing the estimation of any number and combination of SWR data points simultaneously.

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Lucas Totola, Katia Bicalho and Wilian Hisatugu. The initial draft of the manuscript was written by Lucas Totola, and all authors provided feedback on previous versions of the document. Subsequently, all authors reviewed and approved the final version of the manuscript.

**Funding** The authors are grateful to Brazilian agencies CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for financial support.

**Data Availability** Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. The HYBRAS database is available here <http://www.cprm.gov.br/en/Hydrology/HYBRAS-4208.html>.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare.

## References

- Al Majou H, Hassani B, Bruand A (2018) Transferability of continuous- and class-pedotransfer functions to predict water retention properties of semiarid syrian soils. *Soil Use Manag* 34:354–369. <https://doi.org/10.1111/sum.12424>
- Angelaki A, Bota V, Chalkidis I (2023) Estimation of hydraulic parameters from the Soil. *Water Characteristic Curve Sustainability* 15:6714. <https://doi.org/10.3390/su15086714>
- Barros AHC, de Jong Q (2014) Pedotransfer Functions for brazilian soils. *Application of Soil Physics in environmental analyses*. Springer International Publishing, Cham, pp 131–162. [https://doi.org/10.1007/978-3-319-06013-2\\_6](https://doi.org/10.1007/978-3-319-06013-2_6)

- Borgesen CD, Schaap MG (2005) Point and parameter pedotransfer functions for water retention predictions for danish soils. *Geoderma* 127:154–167. <https://doi.org/10.1016/j.geoderma.2004.11.025>
- Botula YD, Cornelis WM, Baert G, Van Ranst E, Congo DR (2012) *Agric Water Manag* 111:1–10. <https://doi.org/10.1016/j.agwat.2012.04.006>
- Botula YD, Nemes A, Mafuka P, Van Ranst E, Cornelis WM (2013) Prediction of Water Retention of Soils from the Humid Tropics by the nonparametric k -Nearest Neighbor Approach. *Vadose Zo J* 12. <https://doi.org/10.2136/vzj2012.0123>
- Botula YD, Van Ranst E, Cornelis WM (2014) Pedotransfer Functions to predict Water Retention for Soils of the Humid Tropic: a review. *R Bras Ci Solo* 38:679–698
- Bouma J (1989) Using Soil Survey Data for quantitative land evaluation. *Advances in Soil Science*. Springer, New York, NY, pp 177–213. [https://doi.org/10.1007/978-1-4612-3532-3\\_4](https://doi.org/10.1007/978-1-4612-3532-3_4)
- Chin KB, Leong EC, Rahardjo H (2010) A simplified method to estimate the soil-water characteristic curve. *Can Geotech J* 47:1382–1400
- Cui YJ (2022) Soil-atmosphere interaction in earth structures. *J Rock Mech Geotech Eng* 14(1):35–49. <https://doi.org/10.1016/j.jrmge.2021.11.004>
- Feuerharmel C, Gehling WYY, Bica AVD (2006) The use of filter-paper and suction-plate methods for determining the soil-water characteristic curve of undisturbed Colluvium Soils. *Geotech Test J* 29:419–425. <https://doi.org/10.1520/GTJ14004>
- Fredlund DG (2006) Unsaturated soil mechanics in Engineering Practice. *J Geotech Geoenvironmental Eng* 132:286–321. [https://doi.org/10.1061/\(ASCE\)10900241\(2006\)132:3\(286\)](https://doi.org/10.1061/(ASCE)10900241(2006)132:3(286))
- Fredlund DG, Fredlund MD (2020) Application of ‘Estimation procedures’ in Unsaturated Soil mechanics. *Geosciences* 10:364. <https://doi.org/10.3390/geosciences10090364>
- Hagan MT, Menhaj MB (1994) Training Feedforward networks with the Marquardt Algorithm. *IEEE Trans Neural Networks* 5:989–993. <https://doi.org/10.1109/72.329697>
- Haghverdi A, Cornelis WM, Ghahraman B (2012) A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. *J Hydrol* 442–443:46–54. <https://doi.org/10.1016/j.jhydrol.2012.03.036>
- Haghverdi A, Öztürk HS, Cornelis WM (2014) Revisiting the pseudo continuous pedotransfer function concept: impact of data quality and data mining method. *Geoderma* 226–227:31–38. <https://doi.org/10.1016/j.geoderma.2014.02.026>
- Haghverdi A, Leib BG, Washington-Allen RA, Ayers PD, Buscher-mohle MJ (2015) High-resolution prediction of soil available water content within the crop root zone. *J Hydrol* 530:167–179. <https://doi.org/10.1016/j.jhydrol.2015.09.061>
- Haghverdi A, Öztürk HS, Durner W (2018) Measurement and estimation of the soil water retention curve using the evaporation method and the pseudo continuous pedotransfer function. *J Hydrol*. <https://doi.org/10.1016/j.jhydrol.2018.06.007>
- Haghverdi A, Öztürk HS, Durner W (2020) Studying Unimodal, Bimodal, PDI and Bimodal-PDI variants of multiple Soil Water Retention Models: II. Evaluation of Parametric Pedotransfer Functions against Direct fits. *Water* 2020 12:896. <https://doi.org/10.3390/w12030896>
- Hodnet MG, Tomasella J (2002) Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer functions developed for tropical soils. *Geoderma* 108:155–180. [https://doi.org/10.1016/S0016-7061\(02\)00105-2](https://doi.org/10.1016/S0016-7061(02)00105-2)
- IUSS Working Group WRB (2015) World reference base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps, vol World Soil Resour Rep 106. FAO, Rome

- Karube D, Kawai K (2001) The role of pore water in the mechanical behavior of unsaturated soils. *Geotech Geol Eng* 19(3):211–241
- Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environ Model Softw* 15:101–124. [https://doi.org/10.1016/S1364-8152\(99\)00007-9](https://doi.org/10.1016/S1364-8152(99)00007-9)
- Maier HR, Jain A, Dandy GC, Sudheer KP (2010) Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environ Model Softw* 25:891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- Masroufi F, Bicalho KV, Kawai K (2008) Laboratory Hydraulic Testing in Unsaturated Soils. *Geotech Geol Eng* 26:691–704. <https://doi.org/10.1007/s10706-008-9202-7>
- Medrado E, Lima JEFW (2014) Development of pedotransfer functions for estimating water retention curve for tropical soils of the Brazilian savanna. *Geoderma Reg* 1:59–66. <https://doi.org/10.1016/j.geodrs.2014.08.003>
- Mermoud A, Xu D (2006) Comparative analysis of three methods to generate soil hydraulic functions Soil Tillage. *Res* 87:89–100. <https://doi.org/10.1016/j.still.2005.02.034>
- Miguel MG, Bonder BH (2012) Soil-water characteristic curves obtained for a Colluvial and Lateritic Soil Profile considering the Macro and Micro Porosity. *Geotech Geol Eng*. <https://doi.org/10.1007/s10706-012-9545-y>
- Minasny B, McBratney AB (2002) The Neuro-m method for fitting neural network Parametric Pedotransfer Functions. *Soil Sci Soc Am J* 66:352–361. <https://doi.org/10.2136/sssaj2002.3520>
- Minasny B, McBratney AB, Bristow KL (1999) Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma* 93:225–253. [https://doi.org/10.1016/S0016-7061\(99\)00061-0](https://doi.org/10.1016/S0016-7061(99)00061-0)
- Montzka C, Herbst M, Weihermüller L, Verhoef A, Vereecken H (2017) Earth Syst Sci Data 9:529–543. <https://doi.org/10.5194/essd-9-529-2017>. A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves
- Nguyen PM, Haghverdi A, de Pue J, Botula YD, Le KV, Waegeman W, Cornelis WM (2017) *Biosyst Eng* 153:12–27. <https://doi.org/10.1016/j.biosystemseng.2016.10.013>. Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils
- Otoni MV, Lopes-Assad MLRC, Pachepsky Y, Rotunno Filho OC (2014) A Hydrophysical database to develop Pedotransfer functions for Brazilian soils: Challenges and Perspectives. In: Teixeira W, Ceddia M, Otoni M, Donnagema G (eds) *Application of Soil Physics in environmental analyses*. Progress in Soil Science. Springer, Cham, pp 467–494. [https://doi.org/10.1007/978-3-319-06013-2\\_20](https://doi.org/10.1007/978-3-319-06013-2_20)
- Otoni MV, Otoni Filho TB, Schaap MG, Lopes-Assad MLRC, Rotunno Filho OC (2018) Hydrophysical Database for Brazilian soils (HYBRAS) and Pedotransfer Functions for Water Retention. *Vadose Zo J* 17:170095. <https://doi.org/10.2136/vzj2017.05.0095>
- Otoni MV, Otoni Filho TB, Lopes-Assad MLRC, Rotunno Filho OC (2019) Pedotransfer functions for saturated hydraulic conductivity using a database with temperate and tropical climate soils. *J Hydrol* 575:1345–1358. <https://doi.org/10.1016/j.jhydrol.2019.05.050>
- Pachepsky YA, Timlin D, Varallyay G (1996) Artificial neural networks to Estimate Soil Water Retention from easily measurable data. *Soil Sci Soc Am J* 60:727–733. <https://doi.org/10.2136/sssaj1996.03615995006000030007x>
- Pham K, Kim D, Yoon Y, Choi H (2019) Analysis of neural network based pedotransfer function for predicting soil water characteristic curve. *Geoderma* 351:92–102. <https://doi.org/10.1016/j.geoderma.2019.05.013>
- Rawls WJ, Gish TJ, Brakensiek DL (1991) Estimating Soil Water Retention from Soil Physical Properties and characteristics. *Communications in Soil Science and Plant Analysis*. Springer, New York, NY, pp 213–234. [https://doi.org/10.1007/978-1-4612-3144-8\\_5](https://doi.org/10.1007/978-1-4612-3144-8_5)
- Reichert JM, Albuquerque JA, Kaiser DR, Reinert DJ, Urach FL, Carlesso R (2009) Estimation of water retention and availability in soils of Rio Grande do Sul. *Rev Bras Ciência do Solo* 33:1547–1560. <https://doi.org/10.1590/S0100-06832009000600004>
- Reichert JM, Albuquerque JA, Solano Peraza JE, da Costa A (2020) Estimating water retention and availability in cultivated soils of southern Brazil. *Geoderma Reg* 21:e00277. <https://doi.org/10.1016/j.geodrs.2020.e00277>
- Ren X, Kang J, Ren J, Chen X, Zhang M (2020) A method for estimating soil water characteristic curve with limited experimental data. *Geoderma* 360:114013. <https://doi.org/10.1016/j.geoderma.2019.114013>
- Saha S, Gu F, Luo X, Lytton RL (2018) Prediction of soil-water characteristic curve for unbound material using Fredlund–Xing equation-based ANN Approach. *J Mater Civ Eng* 30:06018002. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002241](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002241)
- Schaap MG, Bouten W (1996) Modeling water retention curves of sandy soils using neural networks. *Water Resour Res* 32:3033–3040. <https://doi.org/10.1029/96WR02278>
- Schaap MG, Leij FJ, van Genuchten MT (1998) Neural network analysis for hierarchical prediction of Soil Hydraulic Properties. *Soil Sci Soc Am J* 62(4):847–855. <https://doi.org/10.2136/sssaj1998.03615995006200040001x>
- Schaap MG, Leij FJ, van Genuchten MT (2001) Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J Hydrol* 251:163–176. [https://doi.org/10.1016/S0022-1694\(01\)00466-8](https://doi.org/10.1016/S0022-1694(01)00466-8)
- Shahin MA (2013) Artificial Intelligence in Geotechnical Engineering. In: Yang X-S, Gandomi AH, Talatahari S, Alavi AH (eds) *Metaheuristics in Water, Geotechnical and Transport Engineering*. Elsevier, London, pp 169–204. <https://doi.org/10.1016/B978-0-12-398296-4.00008-8>
- Tomasella J, Hodnett M (2004) Pedotransfer functions for tropical soils. *Developments in Soil Science*. Elsevier, pp 415–429. [https://doi.org/10.1016/S0166-2481\(04\)30021-8](https://doi.org/10.1016/S0166-2481(04)30021-8)
- Tomasella J, Hodnett MG, Rossato L (2000) Pedotransfer Functions for the estimation of Soil Water Retention in Brazilian soils. *Soil Sci Soc Am J* 64:327–338. <https://doi.org/10.2136/sssaj2000.641327x>
- Tomasella J, Pachepsky Y, Crestana S, Rawls WJ (2003) Comparison of two techniques to develop Pedotransfer Functions for Water Retention. *Soil Sci Soc Am J* 67:1085–1092. <https://doi.org/10.2136/sssaj2003.1085>
- van Genuchten MT (1980) A closed-form equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Sci Soc Am J* 44:892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Van Looy K, Bouma J, Herbst M, Koestel J, Minasny B, Mishra U, Montzka C, Nemes A, Pachepsky YA, Padian J, Schaap MG, Tóth B, Verhoef A, Vanderborght J, van der Ploeg MJ, Weihermüller L, Zacharias S, Zhang Y, Vereecken H (2017) Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *Rev Geophys* 55:1199–1256. <https://doi.org/10.1002/2017RG000581>
- Vereecken H, Weynants M, Javaux M, Pachepsky Y, Schaap MG, van Genuchten MT (2010) Using Pedotransfer Functions to Estimate the van Genuchten-Mualem Soil Hydraulic Properties: a review. *Vadose Zo J* 9:795–820. <https://doi.org/10.2136/vzj2010.0045>
- Weynants M, Vereecken H, Javaux M (2009) Revisiting Vereecken Pedotransfer Functions: introducing a closed-form



- hydraulic model. *Vadose Zo J* 8:86–95. <https://doi.org/10.2136/vzj2008.0062>
- Wösten JHM, Pachepsky YA, Rawls WJ (2001) Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *J Hydrol* 251:123–150. [https://doi.org/10.1016/S0022-1694\(01\)00464-4](https://doi.org/10.1016/S0022-1694(01)00464-4)
- Yu H, Wilamowski BM (2011) Levenberg-Marquardt training. In: *Intelligent systems*. CRC Press, 12-1-12–16. <https://doi.org/10.1201/9781315218427-12>
- Zhang Y, Schaap MG (2019) 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>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.