Geoderma 446 (2024) 116900

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Estimating the universal scaling of gas diffusion in coarse-textured soils

J. Valdes-Abellan [a,*], D. Benavente [b], B. Ghanbarian [c], P. Moldrup [d], E. Arthur [e], T. Norgaard [e], L. Wollesen de Jonge [e]

[a] Department of Civil Engineering, University of Alicante, Spain
[b] Department of Environment and Earth Sciences, University of Alicante, Spain
[c] Porous Media Research Lab, Department of Geology, Kansas State University (KS), USA
[d] Department of the Built Environment, Aalborg University, Denmark
[e] Department of Agroecology, Faculty of Technical Sciences, Aarhus University, Denmark

ARTICLE INFO

Handling Editor: A. Agnelli

Keywords:
Soil gas diffusion
Soil water retention curve
Van Genuchten model
Percolation theory
Effective medium approximation

ABSTRACT

Gas diffusion, *D*, in partially saturated soils, constitutes a critical topic in soil sciences. However, it is a complex process and this limits its characterization and estimation. In this study, we analyzed and parameterized the soil gas diffusion using a combination of percolation theory (PT) and the effective-medium approximation (EMA). Here, we selected 126 coarse-textured soils with measurements including sand, silt, and clay content, bulk density, organic matter, porosity, soil water content measured at different pressure heads and saturation-dependent gas diffusion. First, we adopted the van Genuchten model, fit it to the soil water retention curve (SWRC), optimized its parameters, and determined the water content at the inflection point. Second, the parameters of the universal scaling law from PT and EMA were optimized by directly fitting the model to the saturation-dependent gas diffusion data. Those parameters are (1) the critical air-filled porosity, ε_c, (2) the crossover air-filled porosity, ε_x, at which the gas movement behavior changes from the percolation theory domain to the EMA domain; and (3) the average pore coordination number, *z*. Next, a multiple linear regression analysis (MLRA) was applied to link ε_c, ε_x and *z* to other soil parameters, such as soil textural and/or hydraulic properties. Uncertainties in our results were evaluated using a jack-knife resampling technique, which involved applying the MLRA more than 7000 times. Results revealed that the most accurate estimations were obtained when both soil textural and hydraulic properties were used simultaneously. However, the use of only soil textural parameters presents practical advantages, as it provides excellent estimations for ε_x and *z*, although not for ε_c. The latter is a critical parameter in the application of the PT and EMA to gas diffusion that requires both the soil basic properties and water saturation curve properties to be correctly estimated.

1. Introduction

Gas transport through soils in particular, or porous material in general, is an important phenomenon. Therefore, an accurate understanding of the phenomenon is crucial for a wide range of applications from carbon sequestration to toxic gas (e.g., radon) movement.

Gas movement in soils comprises the physical processes of gas diffusion and advection and depends on the geometrical and morphological properties of the porous space (Ghanbarian et al., 2018). The same properties also affect any other fluid movement through the porous media like water or oil.

Models explaining gas diffusion, *D*, in variably-saturated porous media can be mainly classified as either empirical or theoretical. The most widely-used empirical models include those of Penman (Penman, 1940a; Penman, 1940b), Millington and Quirk (1961), Jayarathne et al. (2020) and Moldrup et al., (2000a); Moldrup et al. (2001); Moldrup et al., (2000b). In these models, the effective diffusion coefficient decreases with increasing water content in variable-saturated systems. The effective gas diffusion coefficient decreases sharply as soils become saturated with water because large pores become occluded (Scanlon et al., 2001).

Theoretical models, on the other hand, were developed based on the bundle of capillary tubes approach (Burdine, 1953), such as Zheng et al. (2012) and Xiao et al. (2015). Replacing the complex and interconnected pore space of soils with a bundle of non-interconnected tubes is an oversimplification. For further detail and discussion, see Hunt et al.

* Corresponding author.
E-mail address: javier.valdes@ua.es (J. Valdes-Abellan).

https://doi.org/10.1016/j.geoderma.2024.116900
Received 7 November 2023; Received in revised form 20 March 2024; Accepted 27 April 2024

(2013). In contrast, models based on percolation theory (PT) and the effective-medium approximation (EMA) from statistical physics incorporate the geometry and morphology of pore space (Hunt et al., 2014). The PT, originally proposed by Broadbent and Hammersley (1957), describres the properties of connected clusters in a mathematical framework random network, and it determines macroscopic properties of gas diffusion through porous media from its pore-scale connectivity statistics. On the other hand, the EMA theory is a mathematical approximation in which disordered medium is represented by means of an hypothetical ordered medium; properties from the actual medium represent local perturbations of the effective property in the ordered one (Sahimi, 2003; Sahimi, 2011). Ghanbarian et al. (2014) and Ghanbarian et al., (2015a) discussed gas diffusion in percolation clusters and developed a theoretical framework to model gas transport in porous materials and soils. Combining PT and EMA produced a numerical prefactor whose value depends on the probability threshold, p_c, and, p_x, the value above which the behavior of gas diffusion crosses over from the PT to EMA. Finally, the variable p is the probability that a given site (pore body) or bond (pore throat) is active, i.e., filled with the fluid of interest. Those authors showed that above the percolation threshold, the air-filled porosity dependence of the gas diffusion in porous media follows universal scaling. Above the critical air-filled porosity and below the crossover air-filled porosity it conforms to the power-law scaling from the PT with an exponent of 2.0. However, above the crossover air-filled porosity, it follows the linear scaling from the EMA. Both Kirkpatrick (1973) and Keffer et al. (1996) reported a crossover point near 0.75 $cm^3 \cdot cm^{-3}$, consistent with the result of Ghanbarian and Hunt (2014) who analyzed 71 samples with more than 600 measurements. However, Ghanbarian et al., (2015b) demonstrated that the crossover point is non-universal, particularly in homogeneous media such as mono-sized spheres and sand packs.

Following the study of Ghanbarian et al. (2014), Ghanbarian et al., (2015a) proposed the following theoretical model for gas diffusion

$$\frac{D(\varepsilon)}{D_0} = \begin{cases} 0 \quad 0 \leqslant \varepsilon \leqslant \varepsilon_c \\ \frac{\varepsilon_x - 2/z}{1 - 2/z} \cdot \left[\frac{\varepsilon - \varepsilon_c}{\varepsilon_x - \varepsilon_c}\right] \varepsilon_c \leqslant \varepsilon \leqslant \varepsilon_x \\ \frac{\varepsilon - 2/z}{1 - 2/z} \varepsilon_x \leqslant \varepsilon \leqslant 1 \end{cases} \tag{1}$$

where $D(\varepsilon)$ is the diffusion coefficient in the porous medium (also termed as effective gas diffusion coefficient) $[L^2 \cdot T^{-1}]$, D_0 is the diffusion coefficient in the bulk (free space) $[L^2 \cdot T^{-1}]$, z is the average pore coordinator number [-], ε is the air content ($0 < \varepsilon < 1$) $[L^3 \cdot L^{-3}]$, ε_c is the critical air content below which diffusion ceases $[L^3 \cdot L^{-3}]$, and ε_x is the crossover air content at which diffusive transport crosses from the PT universal quadratic scaling to the EMA universal linear scaling $[L^3 \cdot L^{-3}]$.

Characterizing soils and estimating their diffusive and hydraulic properties (e.g., diffusion) are still challenging tasks. The literature on pedotransfer functions is vast and extensive. Numerous pedotransfer functions have been proposed to estimate the water retention curve (Ghanbarian et al., 2015b; Pachepsky and Rawls, 2004; Rawls et al., 2003; Twarakavi et al., 2009) or saturated hydraulic conductivity (Ghanbarian et al., 2017; Schaap et al., 2001; Vereecken et al., 2010). For a review see e.g., Van Looy et al. (2017). Although pedotransfer functions estimating soil hydraulic properties are widely developed in the literature, to the best of our knowledge, the number of studies that proposed pedotransfer functions for gas diffusion parameters estimation is very limited. Numerous pedotransfer functions have been developed and are frequently used for the soil hydraulic parameters (water retention and unsaturated hydraulic conductivity) and the parameter in their well-known constitutive models (especially the van Genuchten models) (Pachepsky and Rawls, 2004; Pachepsky and van Genuchten, 2011; Padarian et al., 2018). Opposite, to our knowledge only very few pedotransfer functions exist for the soil gas phase transport parameters

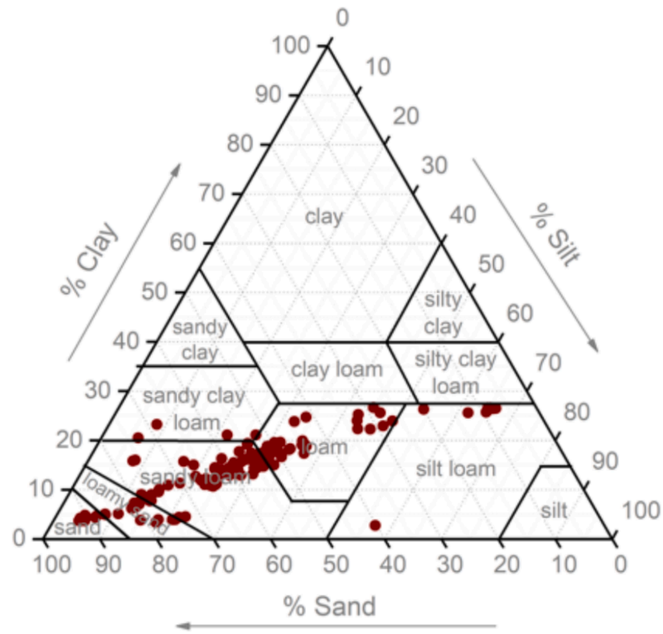

Fig. 1. Soil texture triangle with the location of all 126 soil samples considered in the present study.

(gas diffusivity and air permeability) and the parameters in their constitutive models. These pedotransfer functions are not frequently used since they do not relate to the fundamental soil properties such as soil texture but are typically polynomial functions of soil-air or soil–-water content (Aachib et al., 2004; Van Looy et al., 2017) and, as such having basically the same input, not easier to apply than the more conceptual models like the Water-induced Linear Reduction type models for soil gas diffusivity (Moldrup et al., 2013; Moldrup et al., 2000a). Therefore, the main objective of this study is to develop regression-based relationships to estimate the saturation-dependent gas diffusion model based on the PT and EMA. More specifically, we propose pedotransfer functions for the parameters ε_c, ε_x and z in Eq. (1). In order to achieve this objective and to obtain robust results, using a good quality database featuring laboratory characterization of soil hydraulic properties and soil gas diffusion properties is imperative. In the present study, we collected 126 soil samples, mainly coarse-textured, with available measurements including sand, silt and clay content, bulk density, organic matter, porosity, water retention curve, and saturation-dependent gas diffusion.

2. Materials and methods

2.1. Soil database

The database used in this study includes 189 soil samples. Twenty one samples are from Arthur et al. (2012) and Deepagoda et al. (2013); seventeen samples are from Arthur et al. (2013) and Arthur et al. (2014); four samples from Amoakwah et al. (2017), and the rest are unpublished. For the unpublished data, intact soil cores (100 cm^3) were collected at field capacity moisture content from different field sites in Denmark and then brought to the laboratory for subsequent measurements. For each soil sample the available measurements are sand, silt and clay (%), soil bulk density ($g \cdot cm^{-3}$), organic carbon ($g \cdot 100\ g^{-1}$), and volumetric water content ($cm^3 \cdot cm^{-3}$) obtained at the soil pressure heads of −10, −30, −50, −100, −500 and −1000 hPa; and air-filled porosity ($cm^3 \cdot cm^{-3}$) and relative gas diffusion coefficient, D_p/D_0 (−) obtained at the soil pressure heads of −30, −50, −100, −500 and −1000 hPa. The initial number of soils was later reduced to 126 after cleaning and removing soils lacking some of the analyzed variables. Fig. 1 shows the

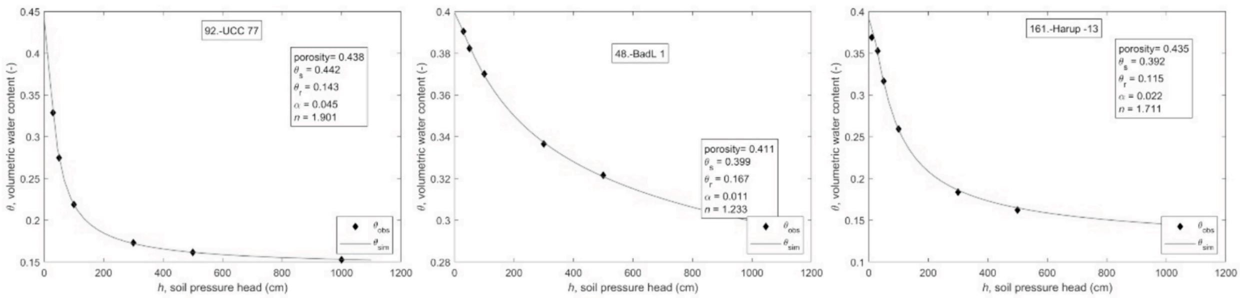

Fig. 2. Three examples of the soil water retention fitted curve to the experimental data.

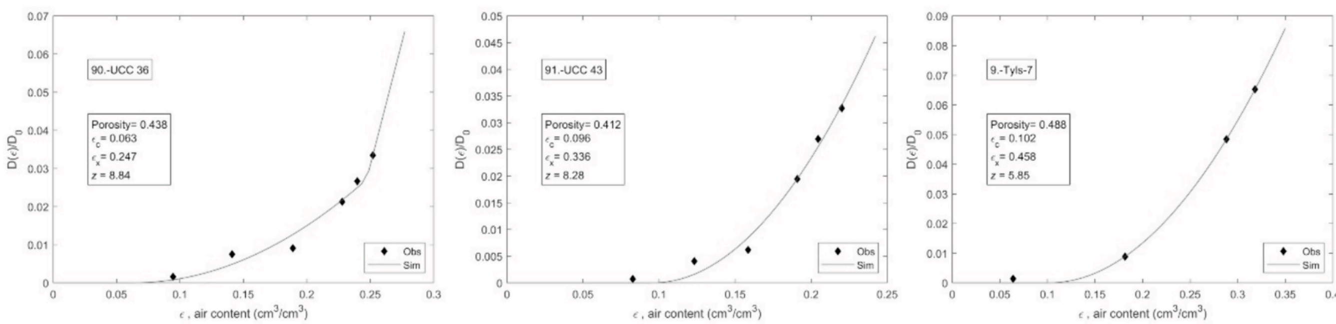

Fig. 3. Three examples of the soil gas diffusion curve fitted to the observed data.

distribution of samples within the soil texture triangle based on the USDA classification. As can be seen, most soil samples are coarse-textured and within the sandy loam and loam soil texture classes.

2.2. Soil water retention curve analysis

The van Genuchten (1980) model (Eq. (2) was fit to the measured soil water retention curve:

$$\theta(h) = \begin{cases} \theta_r + (\theta_s - \theta_r)\cdot[1 + |\alpha\cdot h|^n]^{-(1-1/n)} & \text{h} < 0 \\ \theta_s & \text{h} \geqslant 0 \end{cases} \quad (2)$$

where θ_s is the saturated water content [$L^3 \cdot L^{-3}$], θ_r is the residual water content [$L^3 \cdot L^{-3}$], and α [L^{-1}] and n [-] are empirical coefficients that determine the shape of the water retention curve. In this study, Eq. (2) was fitted to the experimental data and the θ_r, α and n parameters were optimized. The fitting process was assisted by the simplex search method as described by Lagarias et al. (1998). Due to the potential existence of local minimums, we had to replicate the search using different initial points. θ_s was constrained to the [0.9·porosity, porosity] interval. Fig. 2 shows the fit of Eq. (2) for three soil samples. All fits can be found in the Supplementary Material.

The water content at the inflection point, θ_i, was determined after calculating the corresponding pressure head using the following relationship proposed by Dexter (2004) (Eq. (3)

$$h_{in} = \frac{1}{\alpha}\cdot\left(\frac{1}{m}\right)^{1/n} \quad (3)$$

We used a similar approach to fit Eq. (1) to the experimental saturation-dependent gas diffusion data and to optimize the parameters ε_c, ε_x and z. ε_c and ε_x were constrained to the [0, ε_x] and [ε_c, 1] intervals, respectively. Fig. 3 shows the fitted model, Eq. (1), for three soil samples. All fits are given in the Supplementary Materials.

2.3. Multiple linear regression analysis

Multiple linear regression analysis was carried out to establish the relationships linking the parameters of Eq. (1) to other soil properties, such as sand, silt, clay content, bulk density, organic content, porosity, and water retention curves using in-house MATLAB codes. First, we used the SPSS 22.0 software for a preliminary descriptive statistical analysis, scatter diagrams and Pearson's correlation, identifying existing patterns (normal or log-normal) and for quantifying the linear association between two variables. If required, variables were transformed to make their distributions approximately normal before further analysis.

Next, a stepwise multiple linear regression analysis was implemented in MATLAB to develop predictive models for the parameters of Eq. (1) using input variables i.e., the soil textural and hydraulic parameters as well as the water content at the inflection point of the soil water retention curve. The preliminary selection of input variables used to infer the parameters of Eq. (1) were grouped into three input sets, or input scenarios, as follows:

- input set I) basic soil parameters including sand and silt percentages (clay was not considered as it is a complementary value from the two previous), organic carbon content and porosity;
- input set II) SWRC parameters α and n from the van Genuchten model, and soil water content at the inflection point of the SWRC;
- input set III) all previous parameters considered in the input sets 1 and 2.

The aim of this strategy of grouping the input parameters was to determine if the parameters explaining the gas diffusion through soils can be estimated from basic soil properties, from soil hydraulic properties, or a combination of both sets of properties are required to be considered jointly to satisfactorily estimate the soil gas diffusion. The basic input set of parameters selected for (scenario I may vary depending on the research objectives. From our point of view, we consider that all soil parameters included in the scenario I satisfy the requirements of being accessible with very limited experimental capacity and with relatively simple laboratory procedures, so they can be labelled as basic soil parameters.

Parameters ε_c, ε_x and z were considered as the output variables for the stepwise multiple linear regression model. The stepwise multiple

Table 1
Summary of the soil hydraulic properties for the complete database.

Parameter	min	mean (std)	max
Sand (%)	7.19	60.35 (19.89)	91.88
Silt (%)	4.46	25.79 (14.20)	66.27
Clay (%)	2.76	13.86 (6.80)	26.69
OC (%)	0.81	1.68 (0.69)	4.29
Density (g/cm^3)	1.04	1.46 (0.14)	1.73
Porosity (−)	0.35	0.45 (0.05)	0.61
θ_s (−)	0.32	0.42 (0.05)	0.61
θ_r (−)	0.00	0.06 (0.07)	0.24
α	0.01	0.13 (0.24)	1.57
n	1.07	1.39 (0.43)	2.81
θ_i (−)	0.25	0.32 (0.04)	0.49
ε_c	0.00	0.06 (0.04)	0.16
ε_x	0.16	0.32 (0.08)	0.61
z	4.35	10.15 (3.34)	20.00

linear regression not only provides the value of coefficients affecting each input variable but also allows us to obtain the importance of each input variable by the p-value statistic, which is obtained during the fitting process.

In order to determine the uncertainty in the outputs of the stepwise multiple linear regression, a jack-knife resampling technique was applied (Kroll et al., 2015). This process of resampling consists of removing a number of elements from the original dataset and obtaining new subsamples with a smaller number of elements in each of these subsamples. The number of elements to be removed must be small enough to preserve the statistical characteristic of the original dataset, but high enough to obtain a number of subsamples enough to produce robust conclusions. In this study, we decided to remove 2 out of 126 elements from the original complete dataset; by doing so, we were able to get 7875 new subsamples with 124 elements each. The stepwise multiple linear regression analysis was applied to each of these 124-element 7875 subsamples.

After each application of the multiple linear regression model, the goodness-of-fit was assessed by determining the root mean square error (RMSE) and Pearson correlation coefficient (R^2) values. Furthermore, a univariate statistical test was finally applied to identify statistically significant differences in the mean and variance values of RMSE and R^2 for each scenario (i.e.; I-basic soil parameters, II-SWRC parameters and the inflection point, and III-all parameters) and for each output variable (i.e., ε_c, ε_x and z).

Even more, in order to compare the different models (according to each scenario), we obtained the Akaike Information Criterion, after the correction proposed by Hurvich and Tsai (1989) as follows:

$$AIC_c = 2p + n\cdot\log\left(\frac{SSE}{n}\right) + \frac{2p(p+1)}{n-p-1} \quad (4)$$

where p is the number of fitted parameters (i.e.: 4 in scenario 1,3 in scenario 2 and 7 in scenario 3); n is the number of observations and SSE is the sum of square errors after the optimization. As in conventional AIC definition, better values feature smaller values of AIC_C.

3. Results and discussion

3.1. Model fitting to SWRC and gas diffusion data

Results of the basic soil properties and the vG model fitting are given

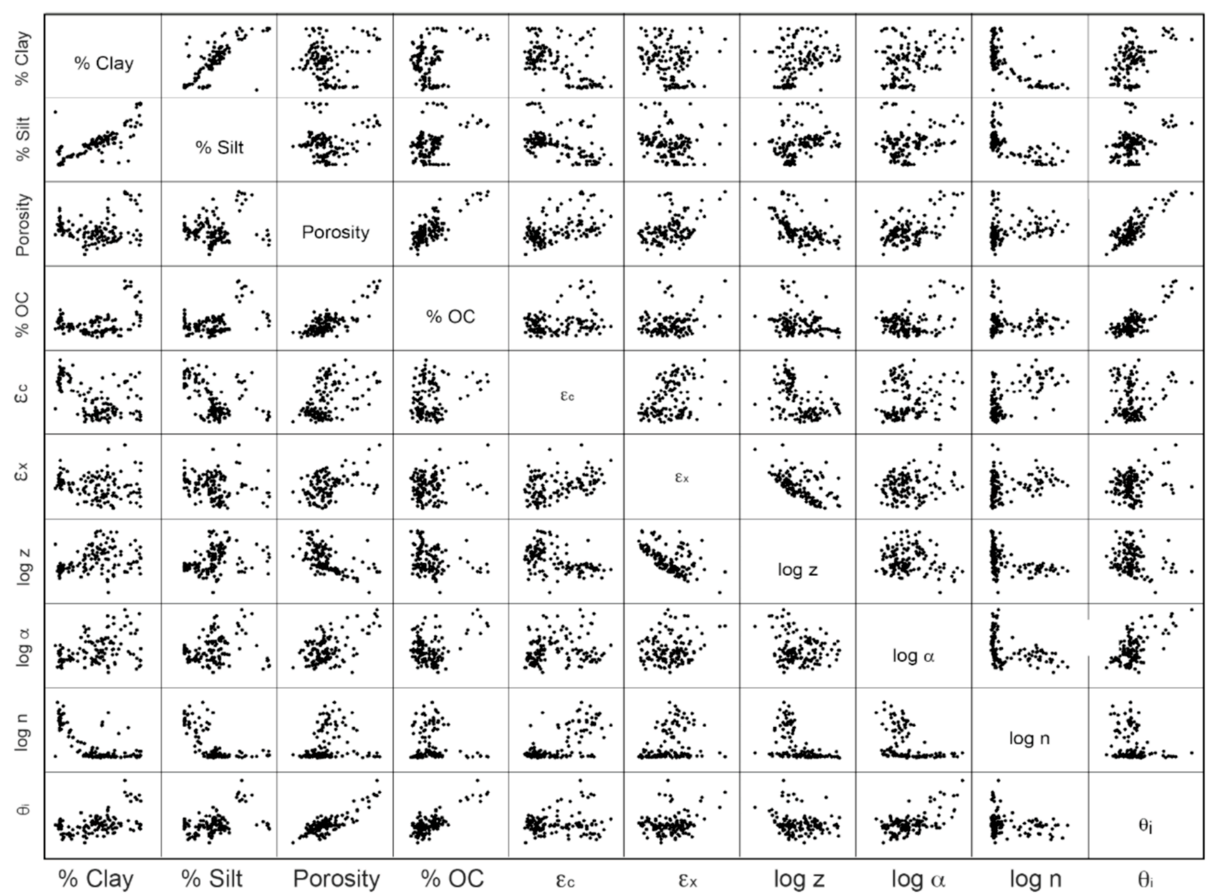

Fig. 4. Correlation matrix of the selected variables: soil parameters (percentages of silt, clay and organic matter, OC, and porosity), effective medium approximation and percolation parameters (air-filled porosity threshold, ε_c, occupation threshold, ε_x, and average coordination number, z); van Genuchten model parameters and the inflection point of the water retention curve, θ_i.

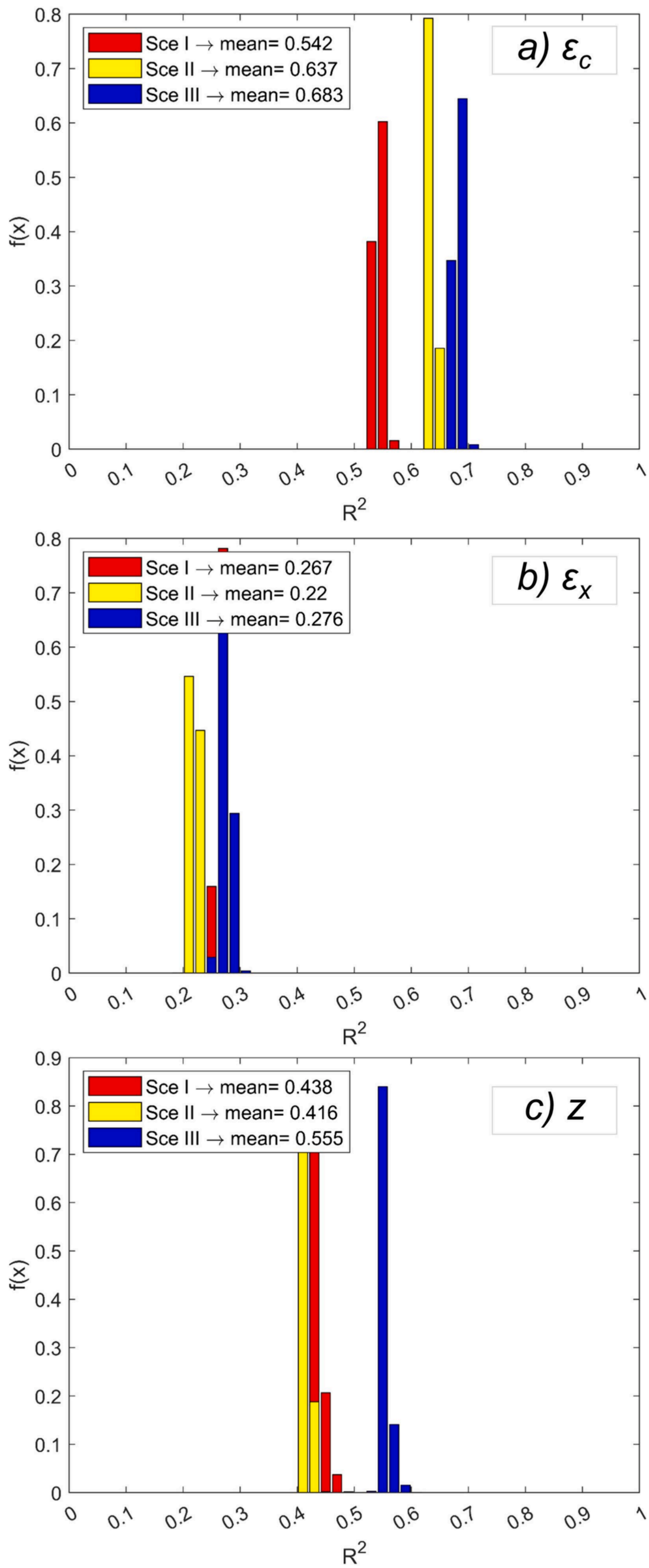

Fig. 5. Histograms of the coefficient of determination (R^2) for the three input data scenarios and for the three output variables (a) ε_c, (b) ε_x and (c) z.

in Table 1. The average value of the parameter n is 1.39 and, as the position of the different quartiles and means reveals, there exists a higher probability of occurrence of n values in the range of 1.1–1.2 (percentile 50th equals 1.14) and a lower number of soil samples with high or very high n values. For the parameter α, we found an average value of 0.13 which is much higher than the median (i.e., 0.04) highlighting the presence of an important number of outliers in the upper part of the graph.

These values are, generally, in accord with studies focused on the development of pedotransfer functions, such as Rosetta (Schaap et al., 2001; Zhang and Schaap, 2017).

We also present the results for the gas diffusion model in Table 1. We found median values of 0.060, 0.316 and 10.154 for ε_c, ε_x and z, respectively. The low values of ε_c demonstrate that the beginning of the diffusion process through the porous media occurs at very low air-filled porosities (i.e., high water content domains). The transition from percolation scaling to the effective-medium approximation (EMA) scaling does not appear in all soils since there exist soils in which the ε_x parameter coincides with the porosity value; however, the general trend for the transition between the two previously mentioned behaviors appear with soil air contents of 70 % of the total porosity.

3.2. Statistics results

Before the analyses between variables, we performed graphical normality tests to identify non-linear trends in the input and output variables, if any. Results showed that α and n followed a log-normal distribution, and consequently, these variables were log-transformed, and their log-transformed values were considered in the regression model. The bivariate analysis identified a linear association between the two variables. Fig. 4 displays the scatter diagrams for a selected set of variables. It is worth noting that this preliminary analysis included the complete dataset (126 elements).

Clay, sand, and silt content were correlated with other parameters (e.g., % Silt had a linear correlation with $R^2 > 0.2$ with ε_c and θ_i, and a non-linear correlation with log α), whereas organic matter did not exhibit any correlation. Bulk density was inversely correlated to porosity to avoid multicollinearity, we removed bulk density and retained porosity in the subsequent statistics analysis.

ε_c did not show a clear linear correlation with any other variable. If any, it seems to be inversely correlated to clay fraction and log(z); a weak direct linear correlation exists between ε_c and porosity and ε_x. Results presented in Fig. 4 also showed a clear inverse correlation between log(z) and ε_x. The same inverse correlation, but less clear is seen between log(z) and porosity.

3.2.1. Regression models

In this study, we considered three scenarios in terms of input setups of multiple regressions to estimate ε_c and ε_x and z, as stated earlier in the methodology section.

Fig. 5 presents the R^2 values obtained from the regression analyses and different output variables. The highest R^2 value was obtained for the parameter ε_c with the highest R^2 value of 0.68 for scenario III. However, we did not find strong relationships linking ε_x to other soil properties. The highest average value of R^2 was 0.28 obtained for scenario III. The average pore coordination number z was quite well estimated using the multiple linear regression model (R^2 equal to 0.56) when using basic and SWRC parameters jointly.

It is worth noting that ε_x was the soil gas parameter worse inferred by other soil basic properties, and on the other hand, the bivariate analysis performed in the previous section to identify linear relations between two variables highlighted a strong inverse relation between ε_x and log (z).

Regarding the comparison between the different scenarios and input data sets, we found that the best fits were obtained for scenario III which includes the basic soil parameters and the SWRC parameters. In our

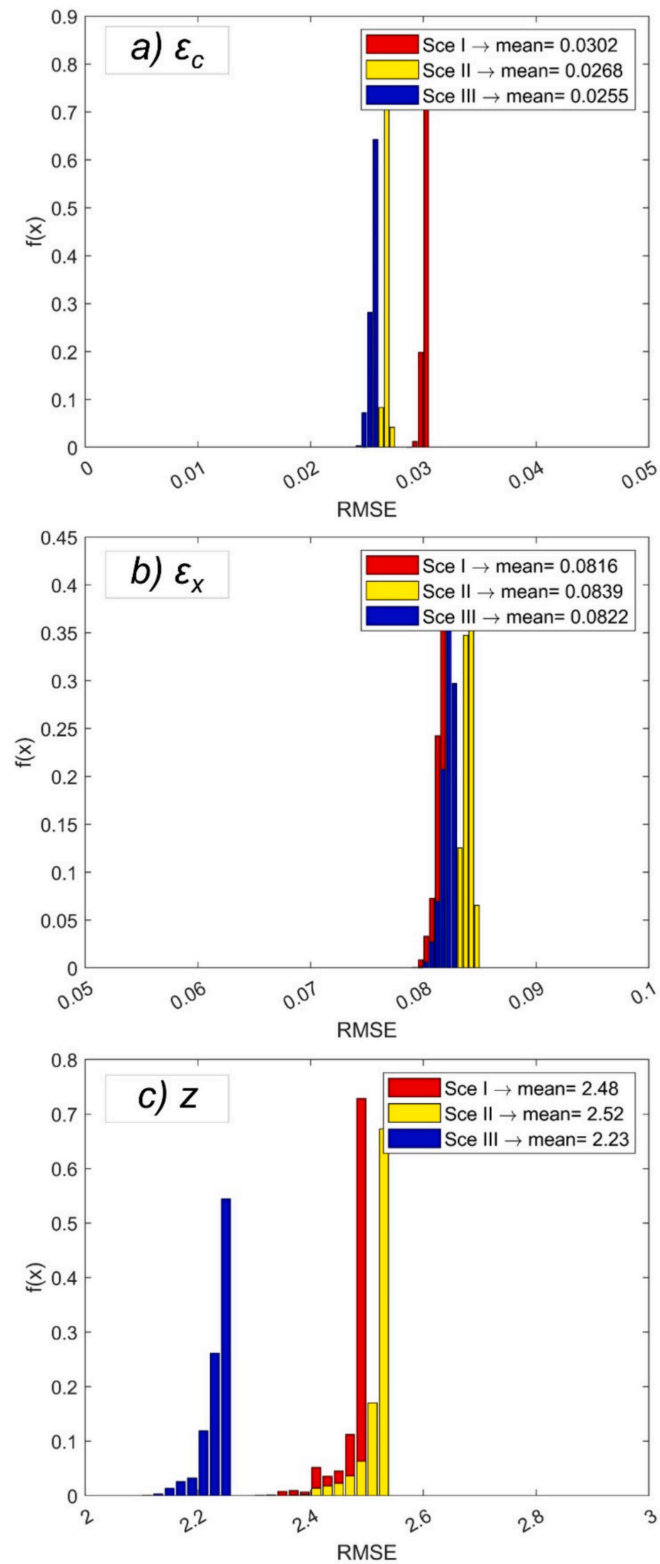

Fig. 6. Histograms of the root mean square error (RMSE) for the three input data scenarios (Sce) and for the three output variables: (a) ε_c, (b) ε_x and (c) z.

comparison of scenarios I and II, we discovered that estimating ε_c solely based on the SWRC parameters yielded significantly better results ($p = 0.05$) than predicting it using the basic soil parameters. Conversely, the prediction of ε_x and z showed slightly higher R^2 values when using basic input soil data. The differences in the mean R^2 values were statistically significant, despite the small disparity in their magnitudes. This fact can

Table 2

AIC_c results for each scenario and soil gas parameter.

	Scenario I	Scenario II	Scenario III
ε_c	−864.8	−895.6	−903.5
ε_x	−618.2	−612.5	−613.1
z	228.3	231.0	206.1

be attributed to the robustness of the results, as they were derived from 7875 analyses.

Differences between scenario III (including all data) and the other two scenarios (i.e., I or II) were noticeable. The *z* parameter was better estimated in scenario III when both basic and SWRC soil parameters and considered jointly as input predictors. This fact did not occur for ε_c or ε_x and results from scenario III are quite close to those obtained for some of scenarios I or II. These conclusions are very similar regardless of the statistic we focus on, RMSE or R^2 (Fig. 6).

Table 2 shows the mean AIC_C statistic for the three scenarios and three soil gas diffusion parameters; a figure depicting the histograms of the AIC_C for the three scenarios has been included in the Supplementary Material section. According to the Akaike criterion, there was no clear superior model for estimating the three soil gas parameters. In the case of ε_c and z, scenario III is the most suitable pedotransfer function, despite the higher number of input parameters; however, for ε_x it was scenario I, which uses as input the basic soil parameters, that provided the most suitable option between the number of parameters and the goodness of fit.

By analyzing the p-value, we obtained insights about which input parameters were the most important ones for estimating the parameters ε_c, ε_x and z. A whisker-box diagram (Fig. 7), corresponding only to the input data scenario III, showed that the input data affecting the soil gas parameter the most was not always the same for the three parameters

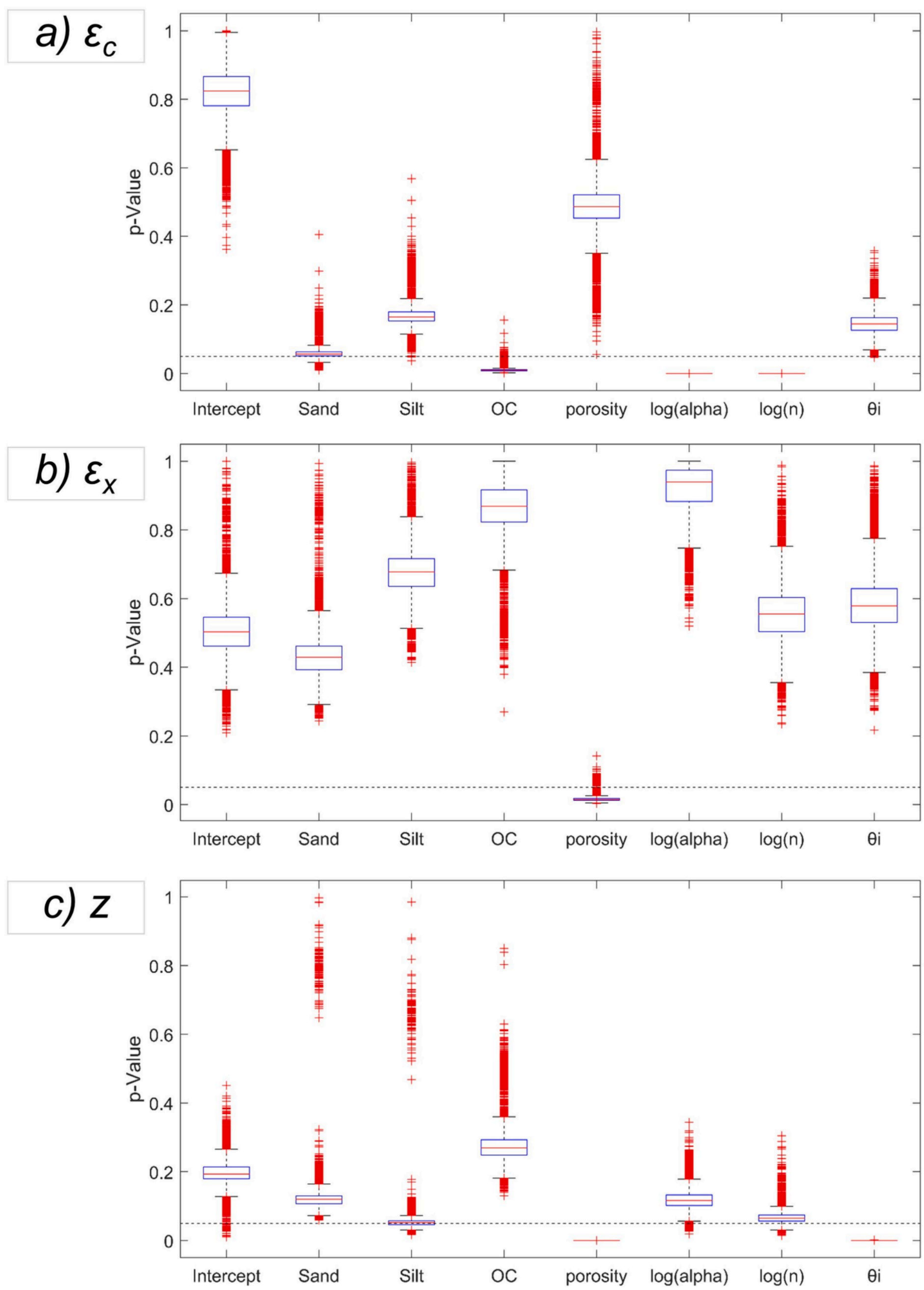

Fig. 7. Whisker-box diagram of the p-value obtained for the input data scenario III and the three output variables: (a) ε_c, (b) ε_x and (c) z.

Table 3
Input coefficients from the multiple linear regression model application for the three output variables ε_c, ε_x and z and for scenarios I, II and III.

ε_c	Input variable	Scenario I	Scenario II	Scenario III
	Intercept	−2.862e-01	5.010e-02	−1.607e-02
	% Sand	2.313e-03		1.159e-03
	% Silt	1.367e-03		1.119e-03
	OC (g·100 g^{-1})	1.054e-02		1.794e-02
	Porosity (−)	3.402e-01		8.897e-02
	log(α) (cm^{-1})		1.750e-02	1.742e-02
	log(n) (−)		1.397e-01	1.083e-01
	θ_i (−)		6.163e-02	−2.307e-01
ε_x	**Input variable**	**Scenario I**	**Scenario II**	**Scenario III**
	Intercept	−2.707e-01	1.597e-01	−1.592e-01
	% Sand	2.607e-03		1.530e-03
	% Silt	1.889e-03		1.058e-03
	OC (g·100g^{-1})	−4.625e-03		−3.505e-03
	Porosity (−)	8.730e-01		9.921e-01
	log(α) (cm^{-1})		1.540e-02	2.758e-04
	log(n) (−)		1.548e-01	3.280e-02
	θ_i (−)		5.022e-01	−2.773e-01
Z	Input variable	Scenario I	Scenario II	Scenario III
	Intercept	2.862e+01	1.112e+01	8.578e+00
	% Sand	−4.979e-02		8.205e-02
	% Silt	2.607e-02		1.366e-01
	OC (g·100g^{-1})	1.959e-01		−6.504e-01
	Porosity (−)	−3.661e+01		−5.723e+01
	log(α) (cm^{-1})		−1.145e+00	−4.812e-01
	log(n) (−)		−8.162e+00	−2.799e+00
	θ_i (−)		−6.101e+00	6.025e+01

(ε_c, ε_x or z). More specifically, we found for ε_c, that porosity and silt percentage variables were the most influential input parameters, while for ε_x and z they were OC and log(α).

Differences in the most important input parameters may be related to the different physical processes that are behind each soil gas parameter. ε_c informs about the point when the soil gas diffusion is initiated (no macroscopic diffusion below it) and so porosity can be a good indicator for this physical phenomenon. ε_x describes the point when the behavior in the transport process crosses from percolation theory's universal quadratic scaling to the EMA's universal linear scaling. Consequently, log(α) becoming the most important predictor, since this is a fundamental pore-network parameter. Finally, z informs about the connectivity or pore coordination, and again log(α) and OC become the most important predictors.

The coefficients affecting each input parameter obtained from the multiple linear regression analysis are shown in Table 3 (these values are the average from the 7875 replications that were applied for each subsample). Using these coefficients, the estimated versus the optimized one is shown in Fig. 8 for each output parameter (Fig. 8 is limited to scenario III). The multiple linear regression model's ability to infer the observed soil gas parameters was different according to which soil gas parameter was considered. The best estimation is obtained for the ε_c parameter.

Although our results are promising and the proposed pedotransfer functions offer a practical way to estimate gas diffusion model parameters from more-basic soil parameters, it should be noted that further improvements are still needed in the estimation of all ε_c, ε_x and z parameters. For instance, further investigations are required using a greater number of measured data points for both soil water retention curves (in the present study we have used only 5–6 observation points for each soils) and saturation-dependent gas diffusion. This limited number of data points may produce an incorrect estimation of soil gas and soil water parameters (see e.g., Fig. 5 in Ghanbarian-Alavijeh and Hunt, 2012). In fact, the literature lacks a large database that includ both soil water and soil gas diffusion data measured over the entire range of water saturation from full saturation to oven dryness.

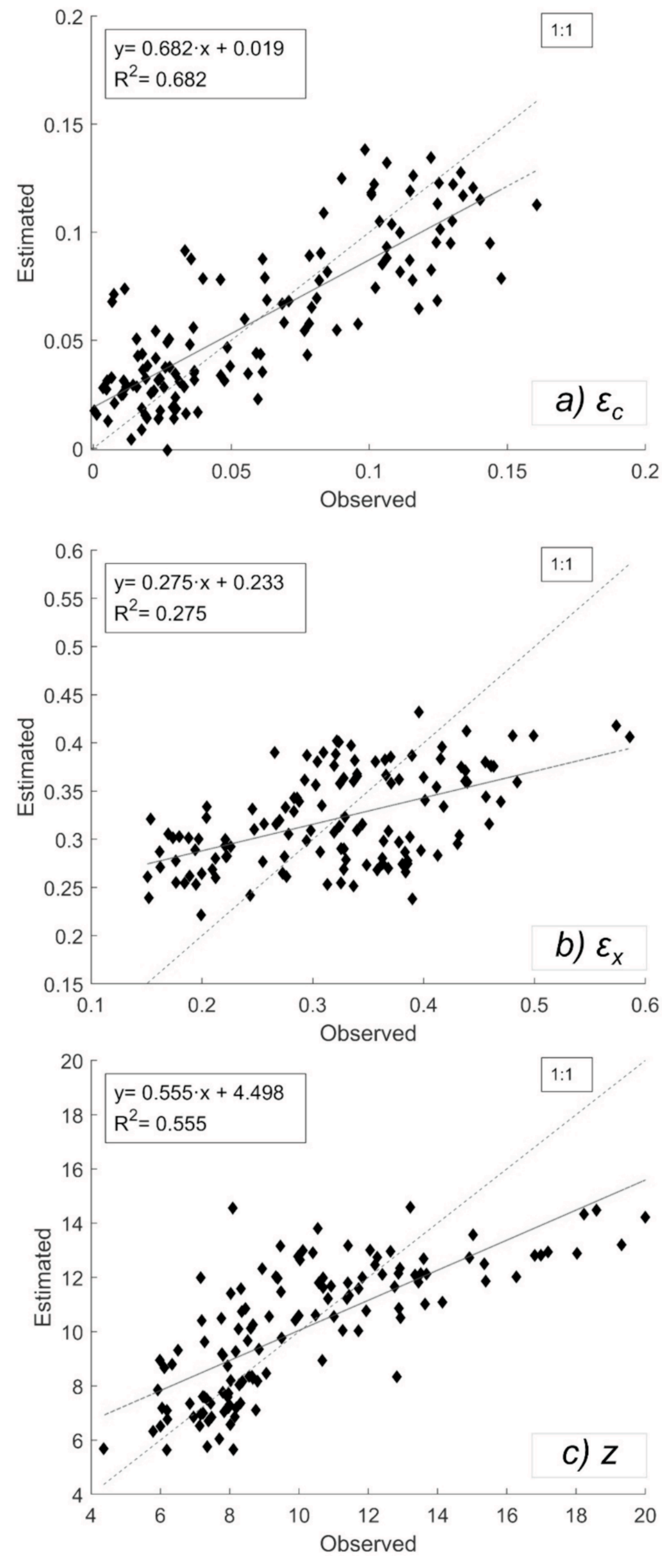

Fig. 8. Estimated versus observed values for the input data scenario III and the three output variables: (a) ε_c, (b) ε_x and (c) z. For each subgraph, the red dashed line is the linear fit to data, and the blue dashed line is the 1:1 line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Conclusions

We set out to examine the relation between some easily obtained soil parameters and the soil gas diffusion parameters ε_c, ε_x and z adopted from the percolation theory (PT) and the effective-medium approximation (EMA). The presented linear regressions constituted a new set of pedotransfer functions to estimate the soil gas diffusion model, Eq. (1), parameters ε_c, ε_x and z from three different levels of input soil parameters: basic soil parameters (scenario I), SWRC parameters (scenario II) and all previous parameters considered simultaneously (scenario III). The effectiveness of the estimation depended on the selected output parameter and the level of input soil parameters.

The strongest linear correlation was observed for the ε_c soil gas diffusion parameter (R^2 equals 0.68) in scenario III. This ε_c relates to the point when the soil gas movement begins; below that threshold, no gas movement exists in the soil porous medium.

The weakest linear correlation was observed for the ε_x soil gas diffusion parameter with a value of the R^2 statistic of 0.27 for scenario III. ε_x determinates the probability (the crossover probability) at which transport crosses from percolation theory's universal quadratic scaling to the EMA's universal linear scaling. The low number of points in each soil sample (between 4 and 6 observation points) may be hindering a correct estimation of this parameter when the soil gas movement changes from one governing process to another.

There is no clear prevalence of any of the scenarios I or II to better infer soil gas parameters. Scenario I (only soil basic parameters) estimated better the ε_x and z parameters, but exhibited a lower performance in estimating ε_c, as indicated by the R^2 statistic values.

While the soil dataset used is indeed valuable, obtaining a greater number of measurements of gas diffusion coefficients for different soil types at various water saturation levels would likely lead to significantly more robust and reliable results.

The outcomes of this study have the potential to aid soil researchers in enhancing the efficiency of soil characterization processes, as well as in fostering additional applications that can stem from well-characterized soil.

CRediT authorship contribution statement

J. Valdes-Abellan: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **D. Benavente:** Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis. **B. Ghanbarian:** Conceptualization, Methodology, Writing – review & editing. **P. Moldrup:** Conceptualization, Methodology, Writing – review & editing. **E. Arthur:** Data curation, Writing – review & editing. **T. Norgaard:** Data curation, Supervision, Writing – review & editing. **L. Wollesen de Jonge:** Data curation, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: David Benavente reports financial support was provided by Spanish Ministry of Science, Innovation, and Universities. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was partially supported by the Spanish Ministry of Science, Innovation, and Universities [grant numbers RTI2018-099052-BI00 and PID2022-139990NB-I00].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2024.116900.

References

Aachib, M., Mbonimpa, M., Aubertin, M., 2004. Measurement and prediction of the oxygen diffusion coefficient in unsaturated media, with applications to soil covers. Water Air Soil Pollut. 156 (1–4), 163–193.
Amoakwah, E., Frimpong, K.A., Arthur, E., 2017. Corn cob biochar improves aggregate characteristics of a tropical sandy loam. Soil Sci. Soc. Am. J. 81 (5), 1054–1063.
Arthur, E., Moldrup, P., Schjønning, P., de Jonge, L.W., 2012. Linking particle and pore size distribution parameters to soil gas transport properties. Soil Sci. Soc. Am. J. 76 (1), 18–27.
Arthur, E., Moldrup, P., Schjønning, P., de Jonge, L.W., 2013. Water retention, gas transport, and pore network complexity during short-term regeneration of soil structure. Soil Sci. Soc. Am. J. 77 (6), 1965–1976.
Arthur, E., Schjønning, P., Moldrup, P., Razzaghi, F., Tuller, M., De Jonge, L.W., 2014. Soil structure and microbial activity dynamics in 20-month field-incubated organic-amended soils. Eur. J. Soil Sci. 65 (2), 218–230.
Broadbent, S.R., Hammersley, J.M., 1957. Percolation processes: I. Crystals and mazes. Math. Proc. Camb. Philos. Soc. 53 (3), 629–641.
Burdine, N.T., 1953. Relative Permeability Calculations from Pore Size Distribution Data. T Am I Min Met Eng 198, 71–78.
Deepagoda, C., Arthur, E., Moldrup, P., Hamamoto, S., Kawamoto, K., Komatsu, T., de Jonge, L.W., 2013. Modeling air permeability in variably saturated soil from two natural clay gradients. Soil Sci. Soc. Am. J. 77 (2), 362–371.
Dexter, A., 2004. Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. Geoderma 120 (3–4), 201–214.
Ghanbarian, B., Hunt, A.G., Ewing, R.P., Skinner, T.E., 2014. Universal scaling of the formation factor in porous media derived by combining percolation and effective medium theories. Geophys. Res. Lett. 41 (11), 3884–3890.
Ghanbarian, B., Daigle, H., Hunt, A.G., Ewing, R.P., Sahimi, M., 2015a. Gas and solute diffusion in partially saturated porous media: Percolation theory and Effective Medium Approximation compared with lattice Boltzmann simulations. J Geophys Res-Sol Ea 120 (1), 182–190.
Ghanbarian, B., Taslimitehrani, V., Dong, G.Z., Pachepsky, Y.A., 2015b. Sample dimensions effect on prediction of soil water retention curve and saturated hydraulic conductivity. J. Hydrol. 528, 127–137.
Ghanbarian, B., Taslimitehrani, V., Pachepsky, Y.A., 2017. Accuracy of sample dimension-dependent pedotransfer functions in estimation of soil saturated hydraulic conductivity. Catena 149, 374–380.
Ghanbarian, B., Hamamoto, S., Kawamoto, K., Sakaki, T., Moldrup, P., Nishimura, T., Komatsu, T., 2018. Saturation-dependent gas transport in sand packs: Experiments and theoretical applications. Adv. Water Resour. 122, 139–147.
Hunt, A.G., Ewing, R.P., Horton, R., 2013. What's Wrong with Soil Physics? Soil Sci. Soc. Am. J. 77 (6), 1877–1887.
Hunt, A., Ewing, R., Ghanbarian, B., 2014. Percolation theory for flow in porous media, 880. Springer.
Hurvich, C.M., Tsai, C.L., 1989. Regression and Time-Series Model Selection in Small Samples. Biometrika 76 (2), 297–307.
Jayarathne, J.R.R.N., Deepagoda, T.K.K.C., Clough, T.J., Nasvi, M.C.M., Thomas, S., Elberling, B., Smits, K., 2020. Gas-Diffusivity based characterization of aggregated agricultural soils. Soil Sci. Soc. Am. J. 84 (2), 387–398.
Keffer, D., McCormick, A.V., Davis, H.T., 1996. Diffusion and percolation on zeolite sorption lattices. J Phys Chem-Us 100 (3), 967–973.
Kirkpatrick, S., 1973. Percolation and Conduction. Rev Mod Phys 45 (4), 574–588.
Kroll, C.N., Croteau, K.E., Vogel, R.M., 2015. Hypothesis tests for hydrologic alteration. J. Hydrol. 530, 117–126.
Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E., 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. 9 (1), 112–147.
Millington, R.J., Quirk, J.P., 1961. Permeability of porous solids. Trans. Faraday Soc. 57, 1200–1207.
Moldrup, P., Olesen, T., Gamst, J., Schjønning, P., Yamaguchi, T., Rolston, D.E., 2000a. Predicting the gas diffusion coefficient in repacked soil: Water-induced linear reduction model. Soil Sci. Soc. Am. J. 64 (5), 1588–1594.
Moldrup, P., Olesen, T., Schjønning, P., Yamaguchi, T., Rolston, D.E., 2000b. Predicting the gas diffusion coefficient in undisturbed soil from soil water characteristics. Soil Sci. Soc. Am. J. 64 (1), 94–100.
Moldrup, P., Olesen, T., Komatsu, T., Schjønning, P., Rolston, D.E., 2001. Tortuosity, diffusivity, and permeability in the soil liquid and gaseous phases. Soil Sci. Soc. Am. J. 65 (3), 613–623.
Moldrup, P., Chamindu Deepagoda, T.K.K., Hamamoto, S., Komatsu, T., Kawamoto, K., Rolston, D.E., de Jonge, L.W., 2013. Structure-dependent water-induced linear reduction model for predicting gas diffusivity and tortuosity in repacked and intact soil. Vadose Zone J. 12 (3).
Pachepsky, Y., Rawls, W.J., 2004. Development of pedotransfer functions in soil hydrology, 30. Elsevier.

Pachepsky, Y.A., van Genuchten, M.T., 2011. Pedotransfer functions. Encyclopedia of Earth Sciences Series. Springer, Netherlands, pp. 556–560.
Padarian, J., Morris, J., Minasny, B., McBratney, A., 2018. Pedotransfer Functions and Soil Inference Systems, pp. 195-220.
Penman, H.L., 1940a. Gas and vapour movements in the soil I. The diffusion of vapours through porous solids. J Agr Sci 30, 437–462.
Penman, H.L., 1940b. Gas and vapour movements in the soil II. The diffusion of carbon dioxide through porous solids. J Agr Sci 30, 570–581.
Rawls, W.J., Pachepsky, Y.A., Ritchie, J.C., Sobecki, T.M., Bloodworth, H., 2003. Effect of soil organic carbon on soil water retention. Geoderma 116 (1–2), 61–76.
Sahimi, M., 2011. Flow and Transport in Porous Media and Fractured Rock: From Classical Methods to Modern Approaches: Second Edition. Flow and Transport in Porous Media and Fractured Rock: From Classical Methods to Modern Approaches. Second Edition.
Sahimi, M., 2003. Heterogeneous Materials I: Linear Transport and Optical Properties.
Scanlon, B.R., Nicot, J.P., Massmann, J.W., 2001. Soil gas movement in unsaturated systems. In: Warrick, A.W. (Ed.), Soils Physics Companion. CRC Taylor & Francis Group, Boca Raton FL (USA).
Schaap, M.G., Leij, F.J., Van Genuchten, M.T., 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251 (3–4), 163–176.
Twarakavi, N.K.C., Šimůnek, J., Schaap, M.G., 2009. Development of Pedotransfer Functions for Estimation of Soil Hydraulic Parameters using Support Vector Machines. Soil Sci. Soc. Am. J. 73 (5), 1443–1452.
van Genuchten, M.T., 1980. Closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44 (5), 892–898.
Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. Rev. Geophys. 55 (4), 1199–1256.
Vereecken, H., Javaux, M., Weynants, M., Pachepsky, Y., Schaap, M.G., Genuchten, V., 2010. Using pedotransfer functions to estimate the van genuchten- mualem soil hydraulic properties: A review. Vadose Zone J. 9 (4), 795–820.
Xiao, B.Q., Fan, J.T., Wang, Z.C., Cai, X., Zhao, X.G., 2015. Fractal Analysis of Gas Diffusion in Porous Nanofibers. Fractals 23 (1).
Zhang, Y., Schaap, M.G., 2017. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). J. Hydrol. 547, 39–53.
Zheng, Q., Yu, B.M., Wang, S.F., Luo, L., 2012. A diffusivity model for gas diffusion through fractal porous media. Chem. Eng. Sci. 68 (1), 650–655.