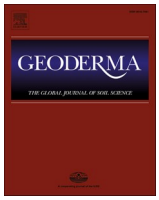
Geoderma 463 (2025) 117550

Contents lists available at ScienceDirect

Geoderma

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



Soil infiltration variability across diverse soil reference groups, textures, and landuse types

Farnaz Sharghi S. [a,*], Sara L. Bauke [a], Mehdi Rahmati [b], Dymphie J. Burger [a], Harry Vereecken [b], Wulf Amelung [a,b]

[a] Institute of Crop Science and Resource Conservation (INRES), Soil Science and Soil Ecology, University of Bonn, Germany
[b] Agrosphere Institute IBG-3, Forschungszentrum Jülich GmbH, Germany

ARTICLE INFO

Handling Editor: Haly Neely

Keywords:
Soil infiltration
Hydraulic conductivity
Variability analysis
World Reference Base for Soil Resources
Hydrological modeling

ABSTRACT

Soil infiltration, a key process in the terrestrial water cycle, is typically measured pointwise, but is often upscaled by averaging across different soil groups or even texture classes, e.g., when parameterizing water movement in land surface models. We hypothesize that for upscaling, in addition to soil texture, infiltration rates/parameters vary also between different reference soil groups and landuse types. Therefore, we analyzed the between- and within-group variabilities of key infiltration parameters, e.g. saturated hydraulic conductivity (K_s) and final infiltration rate (i_c), derived from the Soil Water Infiltration Global (SWIG) database by calculating mutual information and a set of other commonly used statistical measures (e.g., standard deviation) among those classifiers. Results showed that soil texture alone is inadequate to scale up infiltration parameters, leading to lower mutual information and higher standard deviation values of 0.16 and 1.08 for i_c, as well as to 0.16 and 3.65 for K_s, respectively. Similarly, landuse also failed to explain the observed variation in infiltration parameters (with mutual information = 0.28 and 0.14 and standard deviation = 1.10 and 4.08 for i_c and K_s, respectively). In contrast, the World Reference Base soil group was superior to texture and landuse in explaining the observed variability of infiltration parameters, specifically for i_c (with higher mutual information and lower standard deviation of 0.52 and 1.10, respectively). The integrated classification of texture, landuse and reference groups resulted in even higher mutual information and lower standard deviation values (with mutual information values of 0.66 and 0.54 for i_c and K_s, respectively). These results highlight that accounting for the soil classification beyond soil texture should be considered when scaling up the infiltration process.

1. INTRODUCTION

The infiltration of water into the soil is a key process in the terrestrial water cycle (Ma et al., 2016), with important impacts on surface runoff, groundwater recharge, and plant-water uptake, and, thus, on vegetation patterns (Vereecken et al., 2019). Several infiltration parameters such as soil sorptivity (S), saturated hydraulic conductivity (K_s), and initial (i_0) and final (i_c) infiltration rates as well as the decay rate (k_i) at which the infiltration rate decreases from i_0 to i_c, have thus become an essential part in the parametrization of soil water movement in Land Surface Models (LSMs). The parameter S relates to the ability of the soil to absorb water without gravity (Lassabatere et al., 2021). In other words, it quantifies the medium's capacity to absorb water through capillarity. It depends on soil water diffusivity (D), which in turn depends on K_s (Rahmati et al., 2022). The K_s itself, is a metric to measure the ability of the soil to conduct water under the influence of gravity (Rahmati et al., 2020). According to Horton (1941), infiltration dynamics can also be described by an empirical model in which the infiltration rate of water into soil decreases from its initial value (i_0) to a final and constant value (i_c) by considering a decay parameter k.

Soil infiltration, and consequently the associated parameters, are usually measured pointwise, with a range of different instruments. The list of the most common instruments applied for infiltration measurements can be found in the Soil Water Infiltration Global (SWIG) database (Rahmati et al., 2018a). For further applications, e.g., to arrange irrigation systems, to adjust soil and water conservation practices, or to parametrize water movement in LSMs, these pointwise measurements need to be scaled up. Pedotransfer functions (PTFs) have thus been proposed, for instance, to derive K_s from easy to measure parameters such as soil texture (Bouma, 2006; Libohova et al., 2018; Rahmati, 2017;

* Corresponding author.
E-mail addresses: fsha@uni-bonn.de, sharghi.fr@gmail.com (F. Sharghi S.).

https://doi.org/10.1016/j.geoderma.2025.117550
Received 25 May 2025; Received in revised form 30 September 2025; Accepted 8 October 2025
Available online 14 October 2025

Weihermüller et al., 2021). The final upscaling then involves simple interpolation or averaging between different classes of soil texture or landuse types. As an example, existing databases such as Schaap and Leij (1998) and Pachepsky and Park (2015) provided the average K_s for different soil texture groups, which can then be used for upscaling from point measurements to catchment or landscape scale. A more practical example is the use of texture class averaging of soil hydraulic parameters (Carsel and Parrish, 1988) when simulating soil water flow (including the infiltration process) with HYDRUS-1D (Šimůnek et al., 2008; 2016). Upscaling based on soil texture reflects the potential strong relationship between soil hydraulic conductivity, either for the unsaturated state ($K(h)$, where hydraulic conductivity K is a function of the pressure head h in unsaturated conditions) or for the saturated state (K_s), and the relative proportions of sand, clay, and silt (Haghnazari et al., 2015). Accordingly, K_s values are often classified by soil texture, with coarse-textured soils having the highest mean K_s values and fine-textured soils having the lowest (Rahmati et al., 2018a). In addition, i_0 was shown to correlate positively with the content of clay and negatively with silt content (Van Es et al., 1991). Yet, high variabilities of these infiltration parameters, and specifically of K_s within a given texture class remain as a challenge (Rahmati et al., 2018a). As illustrated in Table 1, this variability is larger for clayey than for silty soils, but the large variability of K_s questions the reliability of using solely texture for the upscaling of infiltration parameters. Therefore, it is important to consider other site properties in predicting K_s or other infiltration parameters: de Souza et al. (2019), for instance, showed that infiltration rate depends not only on soil texture, but also on soil porosity, organic matter content, and other environmental factors such as precipitation intensity and landuse. Vereecken et al. (2019), in turn, highlighted that the spatial distribution of infiltration rates is influenced by several soil-related factors, including soil type, local topography and surface cover characteristics.

The Reference Soil Groups in the World Reference Base (WRB) for Soil Resources (Mantel et al., 2023), as well as the soil orders and great groups as defined by the United States Department of Agriculture (USDA) Soil Taxonomy (Soil Survey Staff, 2022), integrate several soil properties into distinct units, which might thus serve as potential classifiers for upscaling soil infiltration parameters. Tejedor et al. (2013) showed, for instance, in a case study in Tenerife (Spain) that Andisols and Entisols, according to USDA classification, had the highest soil infiltration rates, while Vertisols and Aridisols led to the lowest infiltration rates. Soil texture, bulk density (D_b) and water-stable aggregates additionally influenced infiltration behavior, but clay content was most influential, although its importance varied depending on the soil group. Clay had the greatest influence on the infiltration rates in Vertisols, vertic Inceptisols, Alfisols and Aridisols, while its influence was least in Entisols and vitric Andisols; the infiltration rates of the latter two soil groups being rather affected by D_b (Tejedor et al., 2013). Water-stable soil aggregates had the lowest weight in all soil groups except for non-vitric Andisols, where their influence was as large as clay content. It seems thus reasonable to assume that reference soil groups could also be a useful predictor of infiltration parameters at larger scales; yet, to our knowledge, except for the mentioned study by Tejedor et al. (2013) that was restricted to a specific area, a comprehensive, global evaluation of the impact of soil reference group on infiltration rates is still lacking.

Even if soil properties of the reference groups are key for infiltration, soil properties vary with changes in landuse (Fu et al., 2000). The latter are already seen as one of the main factors influencing soil infiltration; yet previous studies were not able to draw general conclusions regarding the impact of landuse change on soil infiltration capacity, due to the complexity of soil–plant systems (Sun et al., 2018). While some authors argue that soil infiltration rate increases with the conversion from forest to agroforestry, e.g., Wang et al. (2015), others report the opposite, e.g., Ma et al. (2007). Dos Santos et al. (2018) showed that both initial and final water infiltration rates and, thus, the total amount of water infiltrating into the soil were larger in the forest areas (natural forest, pine plantation) than in agricultural areas (burned natural rangeland, crop-livestock integration). Zimmermann et al. (2006) found that infiltration capacity decreased with increasing land-use intensity, being lower in pastureland than in plantations, and with slow recovery of infiltration after pasture abandonment. Also, Muñoz et al. (2017) concluded that the variability of infiltration was mainly related to landuse, followed by vegetation type and least related to soil moisture, with all other

Table 1
Variability of K_s values within different soils texture classes (adopted and extended from Rahmati et al., 2018a).

Databases		Rosetta3		Rawls database	Ahuja database	UNSODA database	US Soils K_s data	SWIG database
	(Clapp and Hornberger, 1978)	(Zhang and Schaap, 2017)	(Cosby et al., 1984)	(Schaap and Leij, 1998)			(Pachepsky and Park, 2015)	(Rahmati et al., 2018a)
Parameters Texture class	$logK_s$			$logK_s$/STD(n) log ($cm\ day^{-1}$)				
Sand	1.056	2.81/0.59 (253)	0.82/0.39(−)	2.71/0.51 (97)	3.01/0.45 (82)	2.70/074 (129)	2.17/0.98 (115)	3.02/3.55 (229)
Loamy sand	0.938	2.02/0.64 (167)	0.30/0.51(−)	1.91/0.61 (135)	2.09/0.69 (19)	2.36/0.59 (51)	1.42/0.58 (76)	2.77/3.24 (63)
Sandy loam	0.208	1.58/0.67 (315)	−0.13/0.67 (−)	1.53/0.65 (337)	1.73/0.64 (65)	1.58/0.92 (79)	0.91/0.75 (169)	3.00/3.60 (424)
Silt loam	0.0432	1.28/0.74 (130)	−0.4/0.55(−)	1.04/0.54 (217)	1.24/0.47 (12)	1.48/0.86 (103)	0.15/1.2 (215)	1.85/2.09 (165)
Loam	0.0417	1.09/0.92 (117)	−0.32/0.63 (−)	0.99/0.63 (137)	0.83/0.95 (50)	1.58/0.92 (62)	0.17/1.09 (81)	2.07/2.52 (270)
Sandy clay loam	0.0378	1.14/0.85 (13)	−0.2/0.54(−)	1.29/0.71 (104)	0.81/0.80 (36)	0.99/1.21 (41)	0.17/1.3 (139)	2.11/2.22 (84)
Silty clay loam	0.0102	1.04/0.74 (46)	−0.54/0.61 (−)	0.87/0.55 (47)	1.09/0.78 (21)	1.14/0.85 (21)	−0.21/1.04 (83)	2.47/3.18 (64)
Clay loam	0.0147	0.87/1.11 (58)	−0.46/0.59 (−)	0.67/0.58 (77)	0.79/1.08 (48)	1.84/0.89 (25)	−0.04/1.3 (109)	2.26/2.71 (166)
Sandy clay	0.013	1.06/0.89 (10)	0.01/0.33(−)	1.33/0.33 (9)	−0.03/1.28 (2)	− /−(−)	−1.06/1.68 (21)	− /−(−)
Silty clay	0.0062	0.98/0.58 (14)	−0.72/0.69 (−)	0.82/0.55 (12)	1.15/0.16 (5)	0.92/0.71 (12)	−0.99/1.31 (22)	3.03/3.33 (54)
Clay	0.0077	1.17/0.92 (60)	− /−(−)	0.94/0.31 (34)	1.03/0.83 (31)	1.41/015 (27)	−0.24/0.98 (115)	3.55/4.04 (79)
Silt	–	1.64/0.27 (3)	− /−(−)	1.43/- (3)	− /−(−)	1.75/0.20 (3)	− /−(−)	− /−(−)

STD(n) = standard deviation for the studied sample size n, indicated here behind the $logK_s$ values.

variables ranging in between. Considering the above, it seems therefore logical to consider land-use type also as a potential classifier for the upscaling of soil infiltration parameters, particularly when combined with reference soil groups. Therefore, the main aim of the present study was to evaluate between and within class variabilities of infiltration parameters considering different soil texture classes, WRB soil reference groups and landuse types. We hypothesized that including WRB soil reference groups and landuse will significantly improve the representation of infiltration parameters in relation to the established use of soil texture. In addition, we evaluated to what degree the respective representation of infiltration parameters among different classes depends on the dimensionality (1D or 3D) of the water flow model. To tackle these research questions, we added WRB soil group and landuse data where needed to the comprehensive database of SWIG (Rahmati et al., 2018a), which contains infiltration data collected worldwide, and subjected the infiltration data to variability analysis.

2. MATERIALS AND METHODS

2.1. Study area and dataset

The study has been conducted on a global scale using 5023 individual infiltration curves registered in the SWIG (Global Soil Water Infiltration) database (Rahmati et al., 2018a; Rahmati et al., 2018b). The SWIG database has been developed by collecting in situ infiltration data along with several metadata related to the experimental conditions (initial soil moisture content at the start of the experiments and method used), soil properties, landuse, topography and geographical coordinates of the sites (Rahmati et al., 2018a). To do this research, we first completed SWIG database by embedding WRB reference groups of all data points from SoilGrids database (soilgrids.org) and landuse types from the Global Land Data Assimilation System GLDAS-NOAH (https://disc.gsfc.nasa.gov). Then, soil infiltration curves were fitted over selected infiltration models to obtain infiltration parameters (described in next section). Finally, the variability of infiltration parameters is examined within and between different classes of soil texture, landuse, and WRB reference groups.

We downloaded the WRB reference groups for all geographical coordinates of data points registered in SWIG database using the soilgrids python package (Gan, 2024). Extraction of landuse data for registered geospatial points in SWIG database was done by using Point Sampling Tool in QGIS software (Moyroud and Portet, 2018). Finally, all calculations were made in Octave, a free software licensed under the GNU General Public License (GPL).

SoilGrids is a machine-learning based global soil map derived from $>$ 200,000 profiles and a comprehensive set of environmental covariates, which has been developed across methodological research lineages with uncertainty reporting in the literature. Yet, whilst satellite-based global-scale deep soil property inference appears attractive, independent evaluations have demonstrated region-, depth- and property-specific prediction accuracy levels; categorical output such as WRB reference soil groups can be highly sensitive to local data density and modelling assumptions. In practice, since the SWIG database does not store diagnostic horizon information, we had no way to ground-truth WRB diagnostic horizons in each record, so our analysis treats SoilGrids WRB labels as reliable first-order classifiers and takes local classification errors into account (e.g., Hengl et al., 2017; SoilGrids technical note; independent validations). However, WRB information is thus used only to give some idea at the global scale, but we recommend always checking locally if possible.

2.2. Infiltration models and parameters

We first excluded points with infiltration data measured by rainfall simulator, as well as obtained by linear source and point source methods due to the difficulty of interpreting their data. Linear and point source methods use the advances of runoff on soil surface and the water running out of the slope to determine soil infiltrability (Lei et al., 2006); the linear source method uses a continuous line of sources, while the point source method represents flow from a single point (Mao et al., 2016). After excluding them, therefore, a total of 4178 out of 5035 data points were finally selected from the SWIG database for further analyses. The selected data points were categorized into two groups of 1D (one-dimensional) and 3D (three-dimensional) water flow based on the instruments used to measure the infiltration data. We thus obtained 1D data for 828 data points measured by double-ring infiltrometers, and 3D data for 3350 data points measured by the remaining instruments including single-ring, Guelph permeameter, disc infiltrometer, micro-infiltrometer, mini-infiltrometer, Aardvark permeameter, hood infiltrometer, tension infiltrometer, and the BEST (Beerkan Estimation of Soil Transfer) methods (Rahmati et al., 2018a; Rahmati et al., 2018b).

We used the Horton model (Horton, 1939) to describe the 1D infiltration data, and 3D Haverkamp model (Haverkamp et al., 1994) to describe the 3D infiltration data. It is therefore worth noting that, in this paper, we refer to the Horton model as '1D' and the Haverkamp model as '3D' solely to designate the infiltration experiments on which the data is based, rather than representing the physical form of the infiltration equation. The Horton model (Horton, 1939) reads as follows:

$$i(t) = i_c + (i_0 - i_c)e^{-\kappa t} \tag{1}$$

where $i(t)$ is the infiltration rate [$cm\ h^{-1}$] at time t [h]; i_0 and i_c are the initial and final infiltration rates [$cm\ h^{-1}$], respectively; and κ (h^{-1}) is an empirical constant that reflects the decay in the infiltration rates depending on the soil type, landuse, and management practices. It is worth noting that parameter i_c represents an approximation of field saturated hydraulic conductivity (Horton, 1939).

As stated above, in the case of 3D infiltration data, the quasi-exact implicit formulation developed by Parlange et al. (1982), extended by Smettem et al. (1994), and then redefined by Haverkamp et al. (1994) was applied. This model, mainly known as the Haverkamp model (Moret-Fernández et al., 2020) relies on the following formula:

$$I_{3D} = S\sqrt{t} + \frac{2-\beta}{3}K_{sat}t + \frac{\gamma S^2}{R_D(\theta_f - \theta_i)}t \tag{2}$$

where θ_f and θ_i are the final and initial volumetric water content of soil, β is an integral shape parameter with a default value of 0.6, R_D (L) is the radius of the disc or ring, γ is the proportionality constant with a default value of 0.75 (Haverkamp et al., 1994), and the remaining parameters (S [$cm/h^{1/2}$] and K_s [cm/h]), are defined as described above. Although alternative values of 1.1 (Latorre et al., 2018) have also been reported for β, a default value of 0.6 is commonly used and, therefore, here we also fixed it to 0.6, following the recommendations by Angulo-Jaramillo et al. (2000); Autovino et al. (2024); Haverkamp et al. (1994). The same is true for the γ where 0.75 is the commonly used value (YILMAZ et al., 2023), although some alternative values have been reported for that (YILMAZ et al., 2023). To facilitate the fitting process and to prevent the risk of equifinality (Rahmati et al., 2020) — the principle that the minimum of the objective function can be reached with a wide range of parameter values (Beven and Freer, 2001) —, we approximated the third term to be $A \times t$ with $A = \frac{\gamma S^2}{R_D(\theta_f - \theta_i)}$ being considered as a constant, determined through fitting process. This reduces the number of unconstrained parameters and limits their interaction, an approach which improves parameter identifiability. We also imposed prior constraints for each fitting parameter by setting its range based on physical limits provided in the literature as well as their stability through repeated fits with varying initial values. These steps ensured that the fit to values of parameters had physical meaning while remaining unique.

Briefly, the infiltration parameters including i_0 and i_c and k of the Horton model and S, K_s and A of the Haverkamp model were predicted through fitting of infiltration rate data (I-t sets) to chosen models and

then were subject to variability analysis (see next section). It is worth noting that due to the log-normal distribution of all examined parameters (i_0, i_c, and k in the case of 1D and S, K_s, and A in the case of 3D), we first transformed them into normal distribution using their logarithms and then subjected them to variability analysis. The accuracy of the fittings was evaluated using root mean square error (RMSE), normalized RMSE (NRMSE), coefficient of determination (R^2), and Nash-Sutcliffe criteria (NSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

$$NRMSE = \frac{RMSE}{y_{max}} \times 100 \quad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i-1}^{n}(y_i - \bar{y}_i)^2} \quad (5)$$

where y_i and $\hat{y}_i$, respectively, are the observed and predicted value, or known and estimated values of parameters, $\bar{y}_i$ is the mean of y_i values, y_{max} is the maximum value of y, and n is the number of datasets for each evaluation. The RMSE and NRMSE (normalized RMSE in percent) can be considered as a calculation of the objective function, and values of RMSE and NRMSE close to zero mean a high accuracy. The NSE criterion were used together with the R^2 to assess the quality of the fittings and the accuracy of the predicted values (Rahmati et al., 2019), where a E or R^2 close to unity means higher accuracy.

2.3. Variability analysis

As stated previously, this study aimed to analyze the variability in infiltration parameters across WRB, landuse and soil texture classes to examine which classification is more representative for infiltration parameters. To achieve this, the variability of infiltration parameters within and between different classes of soil texture, landuse, or WRB reference group were analyzed initially using standard deviation (STD) along with the median and mean values of groups. The within group STD measures the distribution of a set of values around the mean (or median). In the context of variability assessments, a low STD indicates that the values of the examined parameter are close to the mean (or median), while a high STD indicates that the values are more dispersed. Therefore, the high within group STD indicates a lower representativeness of the mean or median values for that specific group.

To evaluate between group variability of infiltration parameters, we used the mutual information (MI) analysis, which measures the degree of relatedness between datasets in several ways (Ross, 2014). MI measures the information overlap between two variables. In other words, MI measures how much knowing the value of one variable (for example, soil texture classification) informs about the value of the other variable (for example, K_s). MI can be used as an alternative to the correlation coefficient, as it does not require linearity of the examined variables relationship which is the case for correlation coefficient, and it is applicable between two discrete variables, two continuous variables, or one discrete and one continuous variable (Ross, 2014). In the case of continuous variables, they are initially transformed to discrete variables by applying the nearest neighbor method or the binning method. However, further investigations showed that the nearest neighbor method is more accurate and reliable for estimating MI than the binning method, and leads to consistent and accurate results (Ross, 2014). Also, the nearest neighbor method is stronger when *k-th* is set to a low integer, usually *k-th* = 3 (Ross, 2014). Therefore, we applied the nearest neighbor method to calculate MI between continuous variables of infiltration parameters and discrete categorical variables of classifiers (soil texture, landuse, and/or WRB reference soil group classes). Briefly, for each data point j with a known class, we first determined the *k-th* nearest neighbors (in our case *k-th* = 3) of data point j based on the continuous variable (e.g., K_s) among the data points having the same class as the target data point j (e.g., Regosols in WRB classification). Then, we determined the distance (d) of data point j to its *k-th* neighbor based on the continuous variable: $d = \left|y_{k-th} - y_j\right|$. Then, the number ($m$) of all data points (regardless of their classes) falling in that distance from data point j are counted (see Fig. S14 in Supporting material). Finally, I_j for that data point j was calculated based on the following expression and this was repeated for all individual data points:

$$I_j = \psi(N) - \psi\left(N_{x_j}\right) + \psi(\text{k-th}) - \psi(m) \quad (6)$$

where $\psi(.)$ is the digamma function, N is the total number of data points included in analysis, N_{x_j} is the number of data points having the same categorical class of data point j, and m and *k-th* are defined as above. Finally, when I_j for all data points has been calculated, the average of I_j over all data points determines the MI (Ross, 2014):

$$MI(X,Y) = \langle I_j \rangle = \psi(N) - \left\langle \psi\left(N_{x_j}\right) \right\rangle + \psi(\text{k-th}) - \langle \psi(m) \rangle \quad (7)$$

where the notation $< x >$ implies the average with x referring to I_j, $\psi\left(N_{x_j}\right)$, and $\psi(m)$. An MI value of 1.5 indicates that two groups are identically matched and are very highly correlated. While an MI value close to zero indicates that two groups are completely random and are not correlated at all. Any MI value falling between zero and 1.5 should be interpreted cautiously because the MI value does not signify the strength of the correlation between examined data, it only indicates the probability of stronger correlations. Further clarification of the MI interpretation is provided in the supporting material. Basically, the MI methodology explained above returns only one value among a continuous variable (e.g., K_s) and a classifier (e.g., landuse) and therefore provides information about the between group variability.

Assuming the number of classes in MI estimation is significantly positively biased towards larger values, we adopted a permutation-based significance test to correct for this bias. The class labels for each classifier were randomly permuted 10,000 times and the MI with the continuous variables was recomputed to obtain a null distribution. The observed MI was significantly higher than the highest score obtained under permutation (P-value < 0.0001 in all cases), reflecting high statistical significance that is not due to an increased number of classes alone.

We performed variability analysis for all individual classifiers (e.g. soil texture, landuse and WRB reference groups) as well as the integrated classification. For the integrated classification, we combined soil texture, landuse and WRB reference groups into classes referring to identical textures under identical landuse classes and identical WRB reference groups (e.g. clay texture + cropland + Regosols as an example class of the integrated classification). It is worth noting that in all of the above-mentioned analysis, we only consider those classes that include more than 10 data points, except for combined classifiers among 3D infiltration datapoints, where the number of classes was high and therefore, only classes with more than 50 datapoints were considered for the analysis.

2.4. Modeling exercise

In addition to the variability analysis described above, we performed simple modelling using a random forest (RF) machine learning algorithm to evaluate whether incorporating WRB reference groups and landuse classes into PTFs can improve prediction accuracy for infiltration parameters. Since an extensive modeling exercise was beyond the scope of this paper, we only performed it for K_s, S, and A (in logarithms space), since there was less data available for the 1D infiltration parameters.

We used two different modeling configurations. In the first configuration we used only conventional PTF predictors: clay, silt, sand and

organic carbon (OC) contents and bulk density (D_b). In the second configuration, we included WRB reference groups and landuse classes as additional categorical predictors. To make sure the data was compatible with the RF algorithm, all categorical variables were encoded using One Hot encoding algorithm.

We then compared the accuracy of the developed models between independent test datasets. To do this, we separated our database into two groups of training (80 percent of data) and test (20 percent of data) subsets and trained and tested models using scikit-learn library in python. To conduct a fair modeling exercise, we first removed the outliers from the K_s, S, and A parameters using the interquartile range (IQR) statistical technique before dividing the data into training and evaluation subsets. First, we identified the 25th and 75th quartiles of the K_s, S, and A parameters. Then, we excluded all data points beyond the defined range in following expression from the modeling exercise:

$$IQR = quartile(x, 75\%) - quartile(x, 25\%)$$

$$x < quartile(NSE, 25\%) - 1.5 \times IQR | x > quartile(NSE, 75\%) + 1.5 \times IQR \quad (8)$$

where X stands for either Ks, S, or A.

Model performance was evaluated using the R^2 and root mean squared error (RMSE) for both training and test datasets. It is also worth noting that we conducted a grid search of hyperparameters for tuning and all numerical predictors (e.g., clay, silt, sand and OC and D_b) were normalized using standard scaling algorithm.

3. RESULTS AND DISCUSSION

3.1. Infiltration parameters retrieval

In total, 828 1D and 3328 3D infiltration curves were analyzed in this research. As mentioned in Materials and Methods section, to determine the infiltration parameters, the 1D infiltration data were fitted to the Horton model with i_0, i_c, and k as fitting parameters and the 3D infiltration data were fitted to the Haverkamp model with S, K_s and A as fitting parameters.

In this study, the Horton and Haverkamp equations are used because they are among the most commonly used infiltration equations and have performed well over a wide array of soils and site conditions. Whereas the Horton model delivers an effective empirical characterization of the "concave-to-linear" curve form prevalent throughout the SWIG dataset, the Haverkamp model yields a rational approximating representation consistent with critical hydrological metrics like saturated hydraulic conductivity and sorptivity. While not explicitly representing dual-domain or preferential flow processes — the concern that might be raised by some, prior examination of SWIG data suggests that such behaviors are relatively infrequent (approximately 35 % of cases) and generally linked to exceptional soil properties (Pachepsky and Karahan, 2022). All data points characterized by non-classical infiltration curves were deemed outliers in our analysis (discussed later), which significantly mitigates the threat of biased parameter estimation.

The accuracy of the fittings for both 1D and 3D infiltration curves are illustrated in Fig. S1, provided in supporting materials. The box plot (see Fig. S1-a in supporting materials) shows that the median of the obtained NRMSE value for 1D infiltration data is approximately 0.25 [–] (equal to 25 %) with a median R^2 value of ~ 0.9. These results indicate that most of the measured 1D infiltration data well fitted Horton model with a low NRMSE and a high R^2 indicating an overall good predictive performance. According to NRMSE statistics, the median error of the Horton model's predictions is relatively low compared to the range of the observed data. The R^2 values are generally higher and lie in a narrower range than the NRMSE value, indicating greater consistency in the models' explanation of the variance in the data.

The goodness-of-fit of the 3D infiltration data (see Fig. S1-b in supporting materials) also indicates that the Haverkamp model also fits the data well with the average median NRMSE value (~0.02 [–], equal to 2 %) being very low. The R^2 values are also very high (~0.99) for most of the data points, indicating a better fit of the model to the data. The presence of outliers in both graphs (either for 1D or 3D) indicates variability in model performance, which means that while the models generally fit well to measured data, there are exceptions where the fit is less accurate. Therefore, for the next steps, we discarded the data points with unsatisfactory fitting performance from further analysis for both 1D and 3D infiltration data groups, as it might affect the variability analysis of the fitting parameters. According to the literature (e.g., see Moriasi et al., 2015), an R^2 value of 0.7 is the suggested threshold for satisfactory simulation or fit performance. However, rather than adhering to the subjective choice of 0.7 as the threshold value, we conducted a sensitivity analysis by applying different thresholds for NSE and checking how the MI value of the classification changed (e.g., soil texture, landuse, and WRB). In addition to NSE, we also inspected the NRMSE and R^2 values corresponding to the threshold NSE as supplementary metrics for this analysis. Regarding the 1D infiltration data, the sensitivity analysis yielded threshold values of 0.77 and 17.5 for NSE and NRMSE [%], respectively, and 0.77 for R^2 (see Figs. S2a, S3a, and S4a in the Supporting Materials). However, the 3D infiltration data was not sensitive to such a threshold (see Figs. S2b, S3b, and S4b in the Supporting Materials). Therefore, we used the IQR statistical technique to find the NSE threshold (with $NSE_{threshold} \sim 0.945$) to screen out data points with poor fitting performance, where $NSE_{threshold} = quartile(NSE, 25\%) - 1.5 \times IQR)$ and $IQR = quartile(NSE, 75\%) - quartile(NSE, 25\%)$. This left us with 583 data points for 1D infiltration data and 2908 data points for 3D infiltration data.

There are already numerous papers dealing with the accuracy of infiltration data fittings to various infiltration models (e.g., Clausnitzer et al., 1998; Mirzaee et al., 2014; Van de Genachte et al., 1996). A literature search was conducted to show how our fitting results are compared to the results reported by others. However, it should be noted that almost all research papers evaluate their fitting results by comparing the fitting parameters with their actual values (e.g. fitted K_s with K_s measured in the laboratory). However, Rahmati et al. (2018a) argue that infiltration parameters obtained from the inverse fit of the infiltration models to measured infiltration data mismatch with direct laboratory measurements of for example K_s, since laboratory measurements are usually not representative of field conditions, where factors such as soil structure, macroporosity, landuse, surface crusting, initial moisture variability, and biological activity (e.g., root channels, fauna) significantly influence infiltration behavior. In addition, there is no measured value reported for infiltration parameters used in our analysis (e.g., i_c, i_0, k, and S) and those reported for K_s also are very limited in SWIG database. Therefore, to assess accuracy of the fittings, it might be more reliable if we compare the measured infiltration curves with the fitted infiltration curves rather than comparing the fitting parameters. Thus, we should be cautious when comparing the accuracy of our infiltration curve-fits with the infiltration parameter-fits reported in the literature, which are often done with a limited amount of data. Despite these considerations, we present a comparison of the values of our fit parameters with those from literature as a preliminary insight and providing a general context for the quality of our fits within the broader research landscape.

Mirzaee et al. (2014) evaluated the accuracy of the fitting of eight different infiltration models including Green and Ampt, Philip, SCS (US Soil Conservation Service), Kostiakov, Horton, Swartzendruber, the modified Kostiakov, and the revised modified Kostiakov model to measure infiltration data and showed that all models can fit to their data with accuracy higher than 90 percent. Their results also showed that according to RMSE statistics, the Horton model outperformed the Philip model. Ghorbani Dashtaki et al. (2009) evaluated the fitting accuracy of the Horton model in comparison to Philip and Kostiakov models and showed that the fitting performance of the Horton model is better than

that of Philip and Kostiakov models. The mean R^2 values obtained for the Horton model in the study of Mirzaee et al. (2014) (with 89 observation/infiltration curves) and Ghorbani Dashtaki et al. (2009) (with 125 data points) are reported to be 0.996 and 0.999, respectively, which are comparable to those obtained in our study (with 634 data points), as we observed an R^2 value of 0.913 (Fig. S1).

Rahmati et al. (2019) showed that the Haverkamp 1D model turns out to be more reliable for estimating S and K_s when the infiltration duration is longer. However, one may be careful not to use models beyond their validity range, such as during very short infiltration times, where capillary forces dominate and assumptions of steady, gravity-driven flow are not yet met (Rahmati et al., 2020). The authors also show that the accuracy of the Haverkamp 1D model (which also builds the base for its 3D model, which is used in our analysis) depends on the number of terms included in the model. Comparing the 1-term, 2-term, and 3-term versions of Haverkamp 1D model, the RMSE values obtained for the S were 13.77, 6.47 and 3.68, respectively, and the R^2 values were 0.85, 0.81 and 0.93, respectively. For K_s, the RMSE values obtained for 2-term and 3-term models were 83.2 and 6.3, with corresponding R^2 values of 0.995 and 0.998, respectively (Rahmati et al., 2019). Our results compare well with those obtained by Rahmati et al. (2019). In a recent paper, Vrugt et al. (2024) fitted the Philip model to 1D infiltration data from the SWIG database, indicating a good correlation between measured and estimated S and K_s. A direct comparison between the results obtained by Vrugt et al. (2024) for the Philip model and those obtained by us for the Horton model is not possible as they check the accuracy of the fitting by confronting fitted and measured K_s (with a RMSE value of 35 *cm/h*), while we do that by confronting the measured and fitted infiltration curves (average RMSE = 11.2 *cm*).

3.2. Within group variability of infiltration parameters

3.2.1. Soil texture related variability

In this section, we examine the variabilities of 1D and 3D infiltration parameters. When comparing the variability of i_c, i_0, and k (the 1D infiltration parameters), almost the same patterns across soil texture classes were obtained for the three infiltration parameters. Therefore, in the following, we mainly discuss i_c and provide the results of i_0 and k as supporting material (e.g., see Fig. S5), as i_c occupies an important position compared to i_0 and k, since it is a crucial parameter in modeling infiltration and soil water management (Rice et al., 2014) and also since it is directly correlated with K_s under field conditions. However, it is

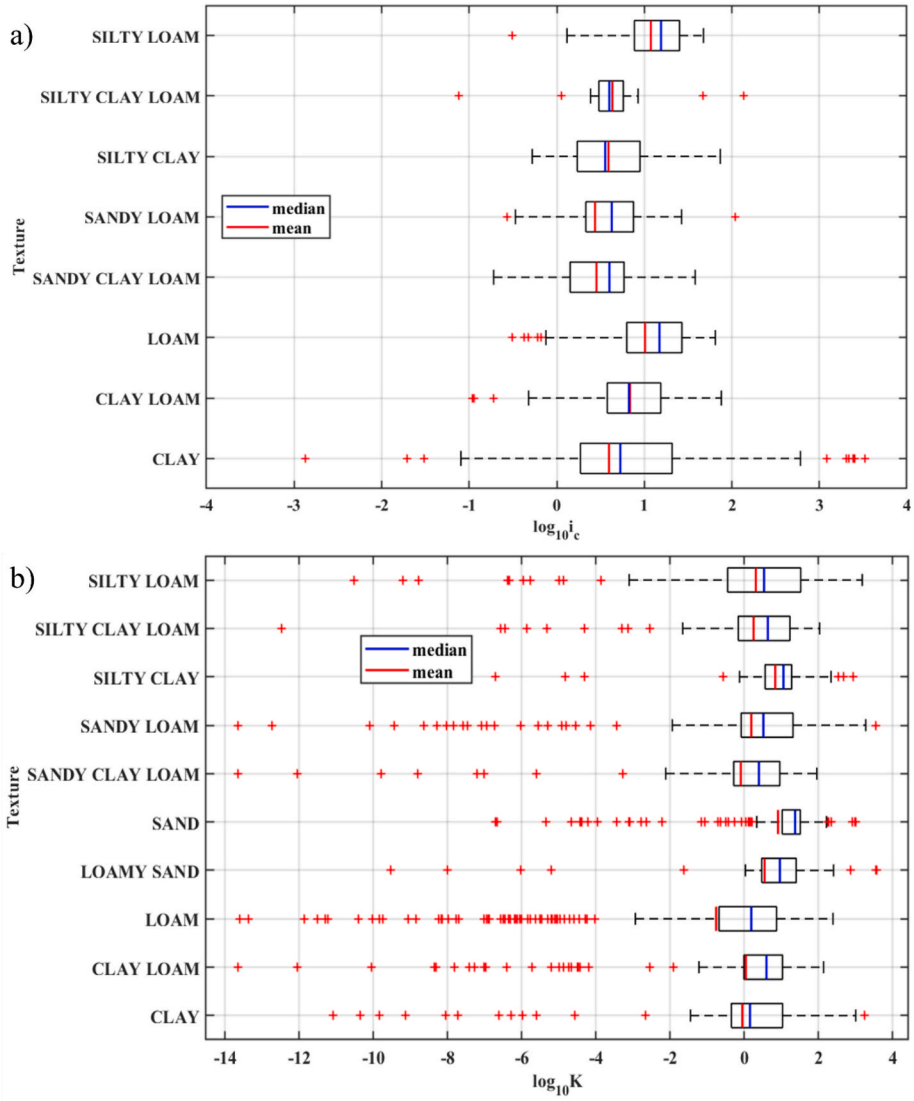


Fig. 1. Boxplots showing the distribution of soil infiltration properties (log-scale) for 1D and 3D infiltration flow across different soil texture classes. a) final infiltration rate (*log10* i_c) in *cm/h* and b) saturated hydraulic conductivity (K_s) in *cm/h*. The blue and red lines represent the median and mean values, respectively, for each soil texture class. Outliers are indicated by red points.

worth noting that although i_c and K_s might be theoretically the same under ideal conditions (Rice et al., 2014), some significant differences between measured K_s and estimated i_c have been demonstrated (Fashi et al., 2010). Therefore, the relationship between i_c and K_s is complex and influenced by several factors, including soil properties, initial soil moisture, and experimental design.

As first impression of the obtained results (Fig. 1 and Table 2), one might notice that the maximum i_c value (Table 2) is obtained for the silt loam texture class with a median value of 1.18 (in log scale, equal to 15.14 cm/h in non-log scale) and the minimum i_c value is obtained for silty clay texture class with a median value of 0.58 (in log scale, equal to 3.80 cm/h in non-log scale). However, also both higher and lower i_c values are obtained for texture classes with dominant silt content (silt loam vs. silty clay). These results indicate how complicated the relationship between infiltration parameters and soil texture classes could be, which should already warn us against relying solely on texture classes when it comes to up-scaling infiltration parameters from point measurements to larger scales.

Fig. 1 illustrates the variability of the 1D infiltration parameters within each soil texture class. According to Fig. 1a, clay-textured soils show the widest range of i_c (in log scale), and have the highest variability among all texture classes, with variation of almost one order of magnitude, including several outliers towards higher rates and some towards lower rates. This is also supported by the highest STD value reported for clay texture class in Table 2. Plotting i_c (in log scale) vs. clay content of soils also confirms the increase of variability in i_c values with the increase of clay content (see Fig. S6 in Supporting Materials) where a greater range of i_c values is obtained for higher clay contents and vice versa. According to Fig. 1a, silty clay loam soils have the lowest variability in i_c (when ignoring outliers) among all texture classes. The remaining soil texture classes fall in between the above-mentioned groups. One important aspect of the obtained boxplots is that the distribution is often asymmetrical for nearly all texture groups, meaning that the average and median value of a specific group are not close to each other. This also suggests that the mean values of the groups are not representative of the group, thus potentially introducing considerable uncertainty when mean values of the different soil texture classes are used for upscaling and modeling purposes.

As it comes to 3D infiltration parameters, similar to the 1D parameters, here we also focus mainly on K_s for discussion, which is a crucial parameter in modeling infiltration and soil water management (Todisco et al., 2023). Without considering the outliers, the 3D K_s parameter showed the lowest variability in coarse-textured soils (e.g., see the shorter whisker box in Fig. 1b obtained for sandy soils) and the highest variability for the loam texture class (Fig. 1b), which is in contrast with the results obtained for 1D i_c parameter, which showed the highest variability in fine-textured soils (e.g., see the large IQR and the longer whisker box in Fig. 1a obtained for clayey soils). Interestingly, the remaining intermediate texture classes (e.g., silt loam, silty clay loam, and sandy loam) also ranked second with the highest variability for K_s. Then, fine-textured soils such as clay ranked third. The remaining texture classes fell between loam and sandy soil in terms of variability but were closer to sandy soil rather than the loam one. According to Fig. 1b and Table 2, neither of the texture classes had a mean value equal (or even close) to the median value, meaning that for none of them the mean value was a good representative of the class, therewith highlighting the need to consider the distribution of infiltration parameters rather than one single value for upscaling water fluxes.

In general, the 3D K_s parameter showed higher variability in all examined texture classes compared to the 1D i_c parameter. This is probably due to the complicated origin of the water flow process in 3D cases. However, it can be argued that the higher variability of 3D K_s parameters between different texture classes compared to the relatively lower variability of 1D i_c parameters can also be due to an additional source of variability introduced by the different types of instruments used to measure 3D infiltration data, whereas this was not the case for 1D infiltration data only measured by double ring infiltrometer. We tested this by repeating the analysis for 3D data, focusing on data measured by each instrument separately (only those with higher number of datapoints, e.g., single ring, disc, mini disc, tension infiltrometers), and results (see Table 3) confirm a significant change in the variability of K_s between different classes. In general, the overall variability of K_s among texture classes obtained by disc and single ring infiltrometers (with overall STD values of 4.74 and 4.89, respectively) are considerably higher than those obtained for mini disc and tension infiltrometer (with

Table 3
Overall variability of 3D saturated hydraulic conductivity (K_s) among different classifiers (soil texture, landuse types, and WRB reference groups) analyzed separated for each infiltration measuring instrument. This table provides the number of observations (Number), mean, median, standard deviation (STD) and mutual information (MI) for 4 infiltrometer types used to measure infiltration rates: Disc, Single Ring, Tension and Mini Disc. The data highlight differences in infiltration behavior between different instruments and shows how site characteristics and measurement methods can influence variation in infiltration.

Category	Instrument	Number	Mean	Median	STD	MI
Texture	Disc	363	−3.64	−0.51	4.74	0.11
	Single Ring	179	−1.96	0.44	4.89	0.16
	Tension	638	−0.10	0.14	2.14	0.31
	Mini Disc	1103	−0.76	0.62	3.46	0.31
	All	2819	−0.98	0.41	3.65	0.16
Landuse	Disc	594	−2.66	0.06	4.59	0.22
	Single Ring	471	−3.00	0.38	5.56	0.06
	Tension	707	−0.31	0.03	2.30	0.10
	Mini Disc	933	−1.05	0.53	3.70	0.51
	All	3278	−1.39	0.29	4.08	0.14
WRB	Disc	589	−2.67	0.06	4.60	0.21
	Single Ring	484	−2.91	0.39	5.54	0.05
	Tension	639	−0.28	0.05	2.41	0.21
	Mini Disc	1122	−0.77	0.61	3.45	0.23
	All	3418	−1.31	0.35	4.07	0.21

Table 2
Variability of final infiltration rates (i_c) in the case of 1D flow and saturated hydraulic conductivity (K_s) in the case of 3D flow among different soil texture classes.

Class£	Number		Mean		Median		STD		MI	
	i_c¥	K_s§	i_c	K_s	i_c	K_s	i_c	K_s	i_c	K_s
Clay	85	211	0.81	−0.20	0.87	0.07	2.06	2.50	–	–
Clay Loam	123	350	0.82	−0.48	0.82	0.56	0.53	3.17	–	–
Loam	77	540	0.99	−2.76	1.16	−0.42	0.58	4.31	–	–
Loamy Sand	–	72	–	−0.82	–	0.87	–	4.13	–	–
Sand	–	282	–	0.74	–	1.30	–	1.79	–	–
Sandy Clay Loam	35	164	0.46	−0.37	0.60	0.35	0.55	2.83	–	–
Sandy Loam	63	515	0.49	−1.09	0.68	0.32	1.30	3.87	–	–
Silty Clay	19	96	0.63	0.62	0.58	1.04	0.56	1.82	–	–
Silty Clay Loam	21	225	0.66	−0.83	0.65	0.43	0.60	3.34	–	–
Silt Loam	75	334	1.02	−1.34	1.18	0.07	0.47	4.08	–	–
All	498	2819	0.79	−0.98	0.84	0.41	1.08	3.65	0.16	0.16

£: all calculations are based on logarithmic bases of i_c and K_s. STD stands for standard deviation and MI stands for mutual information. ¥: 1D flow, §: 3D flow.

overall STD values of 3.46 and 2.14, respectively). Different instrument types, therefore, also act as an additional source of variability in infiltration measurements. This is because different procedures implement various flow regimes and boundary conditions which should be carefully determined. Tension and mini disc infiltrometers, which suppress macropore flow, resulted in less variability (STD = 2.14 and 3.46; Table 3) than single ring infiltrometers that promoted near-saturated flow and can capture more macropore effects (STD = 4.89; Table 3). These methodological effects should be considered in the context of variations of soil classes.

The high variability of K_s in loam soils was in line with the variability reported by Kargas et al. (2021), who also showed that K_s in loam soils exhibit considerable spatial variability with a coefficient of variability of over 70 %. Loamy to silty soils are prone to rapid biopore formation by roots and particularly earthworms (Athmann et al., 2017; Schneider and Don, 2019), which, when present, significantly contribute to water infiltration (Edwards et al., 1988) and, therefore, lead to higher variability of infiltration parameters. Soils with lower K_s tend to have a uniform soil structure in terms of particle size and texture (Ben-Hur et al., 2009). This certainly applies to sandy soils, which show reduced variability in K_s at a higher median. Yet, low variability of K_s was also reported for clayey soils, despite these being prone to drying and swelling, as indicated above when discussing i_c. However, when measured at or near saturation, most of the cracks have already closed, so that K_s might be less impacted by former drying when measured with methodologies suited for the 3D Haverkamp model.

While the A parameter followed a similar pattern to K_s, the pattern is completely reversed for S (see Fig. S7 in supporting materials). This means that contrary to K_s and A, which both exhibit the lowest variability obtained for sandy soils and to some extent for clayey soils, these textures exhibit considerable variability for S, as indicated by the broader IQR (Fig. S7-b in supporting materials) and higher standard deviation values.

3.2.2. Land-use related variability

Considering the extent of 1D infiltration parameters within different landuse classes (Table 4), the highest i_c value was obtained in Deciduous Broadleaf Forest with a median value of 1.30 (in log scale, equal to 19.95 *cm/h* in non-log scale), while the lowest i_c value was observed in Savannas with a median value of −0.05 (in log scale, equal to 0.89 *cm/h* in non-log scale). Croplands showed lower infiltration rates, with mean and median values of 0.08 and 0.24, respectively, which was considerably lower than those for grasslands with mean and median values of 0.98 and 0.87, respectively. This may reflect a disturbance of soil structure and pore continuity in croplands due to ploughing, leading to reduced water permeability in such managed structures (Or et al., 2021; Vereecken et al., 2022). The forests exhibited considerable variability, with mean and median values of 1.24 and 1.30 for Deciduous Broadleaf Forest (among the highest i_c values) and 0.00 and 0.52 for Evergreen Broadleaf Forest (among the lowest i_c values), respectively (Table 4). Fig. 2a shows the within group variabilities of i_c among land-use types. The i_c in croplands and open shrublands ranged by almost two orders of magnitude, indicating the largest variability in log scale. However, when ignoring mixed forests and evergreen needleleaf forests (which have only four and one data points, respectively), Deciduous Broadleaf Forests (with 34 data points) exhibited the narrowest ranges of i_c (in log scale), pointing to limited variability within this land-use type and its respective soils. Grasslands displayed medium variability for i_c but had a notably higher number of outliers compared to other land-use types, thereby increasing the uncertainty associated with its mean value.

In line with the results obtained for the different texture groups, distributions of the i_c (in log scale) within different land-use types were skewed, meaning that the mean values are less representative of classes. Similar results, with only slight differences, were also obtained for i_0 and k among land-use types (see Fig. S8 in supporting materials). A closer inspection of outliers in grasslands (see Fig. 2a) reveals that they correspond to data points around the equator (latitude of 1.15°), which coincide with the presence of Andosols. These points involve double-ring tests, arable land within dominant grassland types with relatively higher organic matter (>2%), and clayey soils with a clay content of > 50 %. These volcanic soils, commonly found in humid regions, develop porous structures that enable high water retention and nutrient storage (Muñoz et al., 2017). However, Andosols are also highly sensitive to environmental changes, such as conversion from natural forests to crop land or shifts in forest vegetation, which contribute to infiltration variability (Tejedor et al., 2013). While the high infiltration rates of ~ 1000 cm/h recorded in Andosols seem extraordinary, a literature review reveals that such rates occur in equatorial regions (e.g., Ecuador), where they have been linked to intact organic mats and large microporosity as well as the use of ring tests which apply large volumes of water (e.g., Suárez et al., 2013). As these measurements are local, method-sensitive upper bounds and not watershed-scale conductivities under or near steady-state conditions, we treat them as outliers in our dataset and refer to them as a special case. As such, we caution against their direct extrapolation in large-scale hydrological models and suggest that future SWIG entries provide detailed metadata on ring method, applied volumes, and vegetation ground cover to allow for better separation of methodological artifacts from true site hydrodynamics.

Among the different land-use types, mixed forests showed the largest variations of K_s (in logarithmic scale; Fig. 2b), accompanied by a large deviation between mean and median values. Although the SWIG database does not include enough data to test this, we postulate that the higher variability of mixed forests likely resulted from macropore-driven flow, which depends on root systems, organic matter

Table 4
Variability of final infiltration rates (i_c) in the case of 1D flow and saturated hydraulic conductivity (K_s) in the case of 3D flow among different Landuse types.

Class£	Number		Mean		Median		STD		MI	
	i_c¥	K_s§	i_c	K_s	i_c	K_s	i_c	K_s	i_c	K_s
Urban and Built-Up	–	18	–	1.7	–	1.58	–	1.14	–	–
Deciduous Broadleaf Forest	34	–	1.24	–	1.3	–	0.24	–	–	–
Barren or Sparsely Vegetated	62	70	1	0.21	1.15	1.4	0.61	3.37	–	–
Evergreen Broadleaf Forest	63	62	0	0.58	0.52	1.2	2.25	2.48	–	–
Open Shrublands	19	326	0.21	−0.6	0.73	0.33	1.62	3.08	–	–
Savannas	26	97	0.04	0.13	−0.05	0.74	0.61	2.85	–	–
Cropland	102	1794	0.08	−1.45	0.24	0.31	0.87	4.24	–	–
Cropland/Natural Vegetation Mosaic	–	113	–	0.52	–	0.55	–	1.41	–	–
Grassland	295	271	0.98	−1.17	0.87	−0.15	0.66	3.14	–	–
Woody Savannas	–	34	–	1.11	–	1.1	–	0.59	–	–
Mixed Forest	–	493	–	−3.3	–	−0.51	–	4.72	–	–
All	601	3278	0.67	−1.39	0.78	0.29	1.1	4.08	0.28	0.14

£: all calculations are based on logarithmic bases of i_c and K_s.
STD indicates standard deviation and MI stands for mutual information,
¥: 1D flow, §: 3D flow.

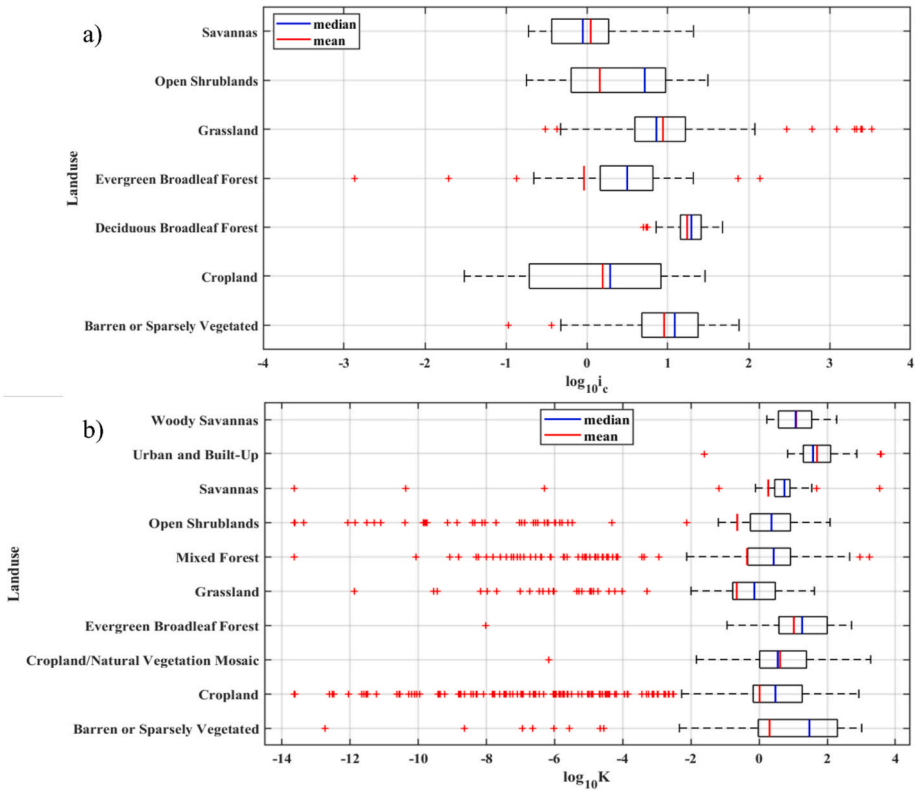


Fig. 2. Boxplots show the distribution of soil infiltration properties (log-scale) for 1D and 3D infiltration parameters across different landuse types. a) final infiltration rate (*log10* i_c) in *cm/h* and b) saturated hydraulic conductivity (K_s) in *cm/h*. The blue and red lines represent the median and mean values, respectively, for each landuse type. Outliers are indicated by red plus signs.

distribution, and soil disturbances, which need to be evaluated in future research. Earlier studies on K_s of mixed forests and grassland soils also indicated high variability due to preferential flow paths in these macropores (Alaoui et al., 2011; Bogner et al., 2010; Smettem, 1987), which enhance infiltration and reduce surface runoff (Alaoui et al., 2011). Similar patterns were observed for the *A* parameter, though the *S* parameter showed different behavior (see Fig. S9 in supporting materials). The *S* parameter exhibited the largest variability for cropland and open shrublands, as reflected by its wide IQR and extended whiskers, whereas variability in mixed forests was minimal. We attribute this to capillary flow processes, which are less influenced by structural heterogeneity in forests (Smettem, 1987). However, in forests with sandy textures or rock fragments, infiltration can be limited, making these areas less favorable for agricultural use. Croplands and sparsely vegetated areas displayed medium variability for K_s (if we exclude outliers). Croplands contained the highest number of outliers of all land-use types. This might result from variations in physical, biological, and management practices. A closer inspection reveals that croplands' outliers belong to data points falling into latitudes of 30 to 54 degrees mainly covered with Cambisols and Luvisols in dry to temperate climate containing the regions with the most croplands on the global scale. Although the lack of detailed data in the SWIG database prevents further investigation, we once again postulate that different management practices and crop types could explain this variability. As mentioned earlier, tillage disturbs pore continuity and can lead to compacted subsoils (Yu et al., 2015), whereas zero tillage can preserve pore continuity but may not necessarily prevent compaction in the long run (Wardak et al., 2022). Possible reasons for increased infiltration rates in vegetated areas include the presence of roots, which improve soil structure, create water channels, and increase porosity (Hao et al., 2020; Liu et al., 2015). Additionally, vegetation increases soil organic matter (OM) content and aggregate stability, which further enhances infiltration (Hao et al., 2020). The extent of these effects varies by vegetation type, with grasses typically performing better than shrubs (Hao et al., 2020; Suárez et al., 2013). Grasslands possess an extensive root system that significantly affects soil structure and infiltration capacity (Mwendera and Saleem, 1997). In many cases, dense root systems at the soil surface can significantly improve water infiltration. Shrubs play a key role in semi-arid ecosystems, particularly on sandy and loamy sandy soils (Marquart et al., 2020). Their root systems create macropores, often accompanied by biopores from burrowing animals such as termites and ants. Furthermore, shrub canopies help to reduce runoff and soil erosion, especially during heavy rainfall events. However, land-use change from grassland, shrubland, or forests to cropland can reduce soil infiltration rates (Sun et al., 2018), which possibly is due to higher distribution of surface soil.

Here 3D infiltration parameters also show higher variability compared to 1D infiltration parameters, which we attribute to a greater variety of measurement instruments used for 3D infiltration data characterization, as discussed previously in texture-related variability section. The results (see Table 3) confirm this, as the extent of variability changes when the analysis of 3D infiltration parameters is restricted to a specific measurement method. It seems that the effect of instruments on landuse related variability of infiltration parameters is even more pronounced compared to texture and WRB related variability, as it is directly related to the different soil surface manipulation activities used for each instrument type. The stronger instrument effects for landuse classes most probably represent the impact of an interplay between the measurement method and land management practices. Single ring infiltrometers presented the highest variability (STD = 5.56), being most sensitive to the heterogeneous surface conditions of saturated flow, while tension infiltrometers had a lowest STD (2.30) as macropores and soil surface features have less influence on tensions. The type of infiltrometer also accounted for substantial variability in the results with

disc and mini disc infiltrometers yielding intermediate values, suggesting that some of the landuse differences could be methodological.

3.2.3. WRB reference group related variability

Table 5 indicates that the Andosols, among different WRB reference groups, stand out as outliers in terms of extraordinarily high infiltration rates under 1D flow condition: the log i_c values even partly exceed 3 (equal to 1000 *cm/h*). Andosols are well known for their porous structure due to the aggregation of allophanes and imogolites, which promote infiltration rates (Woignier et al., 2008). In this regard, isolating the Andosols may truly help in predicting local infiltration properties. Among the other reference soil groups, Ferralsols and Arenosols and to some extent Vertisols exhibit the highest variability in infiltration rates: i_c changed across almost two orders of magnitudes, with the long whiskers indicating the wide scatter of log i_c values within these soil groups (Fig. 3a and Table 5). The high variability of infiltration rates in Vertisols reflects the changes in infiltration with swelling and shrinking behavior of these clay-rich soils (Liu et al., 2010). Dry cracks form temporary channels for water infiltration, contributing to high variability in soil moisture (Návar et al., 2002). A critical crack percentage of 4 % on the surface ensures a considerable increase in infiltration capacity (Cheng et al., 2021). In contrast, in wet periods, swelling closes the cracks unevenly, creating a landscape of micro-highs and micro-lows known as Gilgai microrelief (Kögel-Knabner and Amelung, 2021). The infiltration rate decreases due to the reduced pore space when the cracks close again after wetting, unless heterogeneous saturated flow dominates slow water fluxes. Overall, this soil thus drains poorly when wet. The respective low infiltration rates lead to a large variability in infiltration rates, which depend on soil moisture content and the stage of the wetting/drying cycle. The variability of 3D infiltration parameters in Vertisols is also characterized by the widest range of K_s on a logarithmic scale with long whiskers, which is consistent with the 1D data evaluation (Fig. 3b and Table 5, see also Fig. S10 in supporting materials for S and A). The minimal profile development and pronounced soil heterogeneity due to peloturbation implies high variability in log K_s, which was much larger than that assessed for other soils.

The high oxide contents in Ferralsols frequently form a pseudo-sand structure, which facilitates water infiltration; hence, these soils have a higher average value of i_c than Vertisols (Fig. 3) despite sometimes similar clay contents, as the infiltration pattern frequently creates heterogeneous fingers (Reichenberger et al., 2002). Besides pseudo-sand, pseudo-silt is also formed, and other site properties like soil management, vegetation cover, and faunal activity likely contribute to the high variability of infiltration rates (Basset et al., 2023; Haghnazari et al., 2015). Even if Ferralsols mainly comprise low-activity clays, variations in the contents of clay and organic matter, as well as in remaining quartz grains and in iron and aluminum oxides such as Goethite and Hematite at different sites, can lead to different infiltration rates (Schaefer et al., 2008).

Acrisols exhibited low variability in 3D infiltration rate given the broad range of variability observed for different soil groups. Acrisols have similar chemical properties to the Ferralsols, with acidic pH, pseudosand- and silt-sized aggregation. Yet, Acrisols have a clayey Bt horizon in the subsoil, which can impede infiltration rates. Hence, variations in i_c in the Acrisols were almost as large as that of the Ferralsols, but the mean i_c was lower (Fig. 3a). Ciglasch et al. (2005) reported that water flux in Haplic Acrisols also followed a pronounced fingering, with the diversity of water flux increasing with decreasing seepage flux rate. Above a critical threshold of seepage flux of 2 *mm/d*, the flow pattern and related pesticide transport switched from fingering to matrix-dominated fluxes. Hence, the main reason for variations in infiltration and leaching rates were due to variations in soil moisture content and the intensity of rainfall.

Still, a large variability for the 1D infiltration parameters was found for the Arenosols and Regosols (Fig. 3a, Table 5). Arenosols have a common sandy texture and low organic matter content, which can result in homogeneous infiltration. However, the overall water-binding capacity is low due to their coarser texture, leading to less variability in moisture content between different samples. This contributes to the shorter whiskers (less variability in infiltration parameters) by almost one order of magnitude compared with those of the Ferralsols and Vertisols (Fig. 3a). In contrast, Arenosols and Regosols frequently have sparse and patchy vegetation, which leads to heterogeneous infiltration at a given site. Arenosols are also common soils on former dunes, where the landscape is not necessarily flat. Regosols, in addition, can be the remains of eroded sites and are often located at sloped positions (Stürmer et al., 2009). The orientation and steepness of the slope influence the rates of infiltration. Additionally, Stürmer et al. (2009) indicated that water infiltration in the Regosols can be influenced by saprolite fracture configurations, granulometry, relief conditions, and landuse (Stürmer et al., 2009). When even stones are present, their size, shape, and orientation significantly influence infiltration parameters (Ma and Shao, 2008). Larger stones have been observed to obstruct the infiltration process more thoroughly than smaller stones, with spherical

Table 5
Variability of final infiltration rates (i_c) in the case of 1D flow and saturated hydraulic conductivity (K_s) in the case of 3D flow for different soil groups.

Class£	Number		Mean		Median		STD		MI	
	i_c¥	K_s§	i_c	K_s	i_c	K_s	i_c	K_s	i_c	K_s
Andosols	17	–	2.98	–	3.09	–	0.49	–	–	–
Fluvisols	–	58	–	−0.76	–	1.6	–	5.29	–	–
Chernozems	15	45	1.38	−3.9	1.51	−1.09	0.44	4.99	–	–
Alisols	27	–	1.26	–	1.33	–	0.24	–	–	–
Arenosols	11	244	−0.39	0.89	−0.03	1.31	1.93	1.21	–	–
Leptosols	59	150	1.07	−1.44	1.17	0.25	0.52	3.7	–	–
Calcisols	130	149	0.99	0.48	1.05	1	0.41	2.35	–	–
Luvisols	15	1030	0.9	−1.96	0.95	−0.05	0.45	4.32	–	–
Acrisols	89	39	0.38	1	0.48	0.94	0.6	0.68	–	–
Ferralsols	24	99	−0.47	0.49	0.54	0.93	3.05	2.86	–	–
Kastanozems	–	40	–	0.31	–	0.92	–	2.25	–	–
Gleysols	–	30	–	0.7	–	0.75	–	0.21	–	–
Cambisols	105	1312	0.69	−1.69	0.73	0.93	0.37	4.41	–	–
Regosols	23	162	0.56	−0.81	0.67	−0.26	0.57	2.19	–	–
Vertisols	15	20	0.32	−2.21	0.56	0.12	0.93	3.88	–	–
Lixisols	–	17	–	−0.27	–	0.46	–	3.46	–	–
Podzols	32	23	−0.97	−0.88	−0.97	0.28	0.28	3.23	–	–
All	562	3418	0.69	−1.31	0.79	0.35	1.06	4.07	0.53	0.21
All*	527	3075	0.87	−0.88	0.96	0.45	0.48	3.15	0.52	0.14

£: all calculations are based on logarithmic bases of i_c and K_s. STD stands for standard deviation and MI stands for mutual information. ¥: 1D flow, §: 3D flow, *: all after excluding Arenosols and Ferralsols as outliers with higher STD values for i_c.

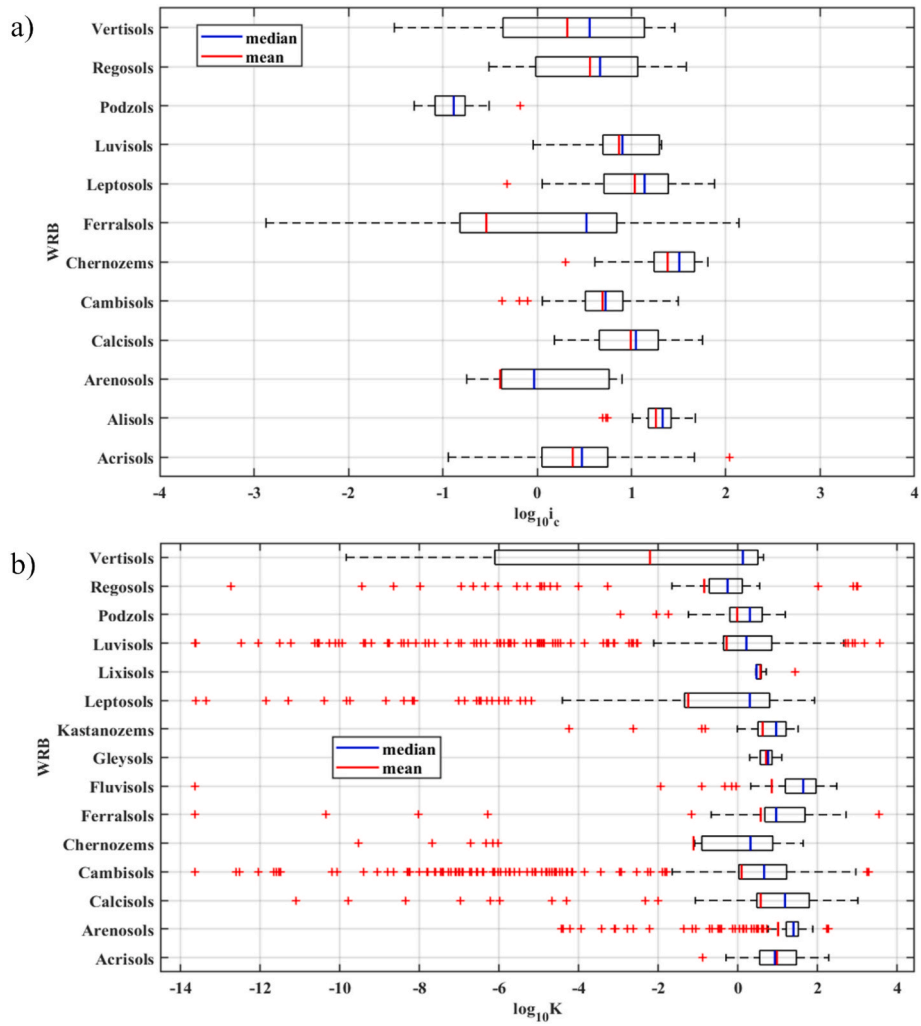


Fig. 3. Boxplots showing the distribution of soil infiltration properties (log-scale) for 1D and 3D infiltration parameters across different WRB reference groups. a) final infiltration rate (*log10* i_c) in *cm/h* and b) saturated hydraulic conductivity (K_s) in *cm/h*. The blue and red lines represent the median and mean values, respectively, for each WRB reference group. Outliers are indicated by red plus signs.

stones increasing levels of infiltration compared to other shapes (Ma and Shao, 2008). Yet very stony soils are frequently no longer assigned to Regosols but to Leptosols.

Regarding the variability of K_s, Regosols, Calcisols, and Arenosols exhibited low variability in K_s. The mean and median values for these soil types are almost close to each other, indicating a relatively symmetrical distribution of the measured properties (hydraulic conductivity, porosity, and Db). Low variability reflects the relatively uniform texture and large pore structure of Arenosols, which are commonly sandy in nature. Sandy soils allow water to flow consistently under saturated conditions, reducing variability in K_s. Regosols are frequently sandy, whereas the potential presence of stones likely affected infiltration in Calcisols less clearly in this data subset than it possibly did in the 1D infiltration dataset. The amount of data for Kastanozems, Solonetzes, Gleysols, and Fluvisols is still small, so we refrain from defining soil group-specific infiltration patterns for these soils due to a smaller number of data points. For the other soils, Podzols and Alisols exhibit the lowest variability in infiltration parameters (Fig. 3b and Table 5). Podzols usually form on sandy parent material, which facilitates infiltration and thus promotes i_c (Fig. 3a). Besides, Podzols are common in forest soils, especially in needleleaf and broadleaf forests, reflecting an interaction between soil properties and the effects of landuse (Törmänen et al., 2020). The low variability of infiltration rates in Podzols likely reflects their uniform sandy to loamy sand texture with predominant quartz composition, thin organic A horizons, and stable management conditions. The remaining variation in i_c may be due to variations in the density of the Bs horizon and, similar to Acrisols, in variations of fingering of the infiltration with variable moisture contents.

Alisols usually have a well-developed illuvial clay subsoil horizon (Shoji et al., 1982), which reduces water infiltration, but on the other hand likely also reduces the variability of infiltration rates. The presence of 1:1 clay mineral, dominance of aluminum, and the acidic environment stabilizes the soil structure, thus contributing to uniform infiltration rates across different profiles (Fig. 3a), although this argument could also apply to Acrisols, which showed larger variability in i_c. Possibly, the occurrence of Acrisols in more humid climates, sometimes with more than 2000 mm rainfall, additionally affects the related variability in infiltration rates.

Regarding the within-group variability of i_c, Table 5 shows that Arenosols and Ferralsols have exceptionally higher STD values (1.95 and 3.05, respectively) compared to others (smaller than 0.93 and mostly lower than 0.5), leading to higher mean STD values for all WRB reference groups. Excluding them leads to a lower overall mean value of STD (with a mean STD of 0.48 for i_c), which is much lower than those obtained for texture classes (with an overall mean STD of 1.08) and landuse types (with an overall mean STD of 1.1). This already confirms our hypothesis that the WRB reference group classification, when used to upscale pointwise measured infiltration parameters, can be much

more representative than soil texture and landuse type classifications. Even without excluding these outliers, the mean STD value obtained is still slightly lower than those for soil texture and landuse types.

In general, silty or clayey soils with higher organic matter contents, like Chernozems, Humic Cambisols, Nitisols, and the Luvisols, have higher water retention and availability than the more sandy Arenosols (Costa et al., 2013). These soils, rich in carbon, exhibited moderate variability in infiltration rates (Fig. 3). The standard deviation values of i_c for different soil types also support this aspect (Table 5). Noteworthy, the results for different soil types show that for almost none of the soil types, the median value is near the center of the overall distribution of i_c values (Fig. 3a and Table 5). This also applies to two other parameters, i_0 and k, and thus, we refrain from elaborating on them for the sake of brevity (see Fig. S11 in supporting materials).

It is also worth noting that clear instrument effects were also evident in the WRB reference groups variability analysis of the 3D infiltration parameters (Table 3). Tension and mini disc infiltrometers produced less variability (with STD = 2.41, and 3.45) than single ring infiltrometer operated under near-saturated conditions (with STD 5.54). Greater variability of single ring infiltrometer probably results from the larger influence of macropore/micro-structure, cracks, and structural heterogeneity, all of which vary markedly among soil groups. By design, tension-based methods minimize macropore flow contributions and provide more consistent measurements among WRB classes. This suggests that some of the variations in what has been contrasted as WRB reference groups are probably exaggerated by method effects and hence highlights the importance of correcting for instrument biases when comparing infiltration characteristics from one soil group with another.

3.3. Between group variability of infiltration parameters

We calculated MI for different infiltration parameters including 1D (i_c, i_0, and k) and 3D (S, K_s, and A) flow conditions with respect to different classifiers (soil texture classes, landuse types, WRB reference groups). As mentioned before, the purpose of calculating the MI was to determine which classifier (e.g., landuse types, soil texture classes, or WRB reference groups) contributes more to explaining the variability of the target variables (e.g., i_c or K_s). We can consider three conditions to interpret the MI results with a valid range of 0–1.5 (see supporting materials): Higher MI values (closer to 1.5) indicate a strong relationship between the classifier (e.g., WRB) and the target variable (e.g., i_c), and thus a greater contribution from classifier to explain i_c variability, i.e., the classifier is more representative to upscale the parameter. A low MI indicates a weak relationship between classifier and target variable, meaning that knowing the classifier provides less information about the target variable. An MI value of zero indicates that the classifier and target variable are completely independent and knowledge of classifier provides no information about target variable.

The MI values obtained between different classifiers and i_c indicate that WRB (with a MI value of 0.52) contributes most to explaining the variability of the i_c, followed by landuse (with a MI value of 0.27) and soil texture (with a MI value of 0.15) (Table 6). This basically means that the soil texture is not representative enough for infiltration parameters and more specifically for i_c, in contrast to common assumptions in literature. The results also show that soil texture hardly informs about i_0 and k due to low MI values of 0.08 and 0.10, respectively (Table 6). In contrast, the WRB reference soil groups provided more informative infiltration parameters, showing MI values of 0.52 for i_c, 0.38 for i_0, and 0.30 for k, which are more than twice/three times higher than the MI values between those parameters and soil texture. Although we cannot argue that this also means that the strength of the relationship between infiltration parameters and WRB reference soil groups is twice (or more) as high as that with soil texture, we can certainly claim that WRB reference soil groups are two times more informative about soil infiltration parameters and consequently, WRB reference soil groups can also be the more representative predictors of infiltration parameters than soil texture classes. This is not a big surprise as the WRB classification system had already proven its validity in supporting soil management practices particularly for addressing water management, salinity and sodicity issues (Naidu and Rengasamy, 1993). The WRB system is organized around diagnostic horizons (and properties or materials) that represent soil-forming process and conditions (Schad, 2016). This broad classification considers various soil parameters including but not limited to texture, OM, and D_b (Salehi, 2018; Schad, 2016). However, it should be noted that these diagnostic horizons are usually not at the soil surface, whereas infiltration is measured at the soil surface. One might then ask why diagnostic subsoil horizons explain so much of the variability in surface infiltration. Addressing this question requires further research, which will be pursued in future work.

Table 6
Mutual Information Among Different Infiltration Parameters Including 1D (logarithms of i_c, i_0, and k) and 3D (logarithms of S, K_s, and A) and Soil Classifiers (Soil Texture Class, LandUse Types, WRB Reference Soil Groups).

Flow type	Parameters	Soil texture classes	Landuse types	WRB soil groups
1D	*Log* i_c	0.15	0.27	0.52
	Log i_0	0.08	0.20	0.38
	Log k	0.10	0.17	0.30
3D	*Log* S	0.20	0.18	0.26
	Log K_s	0.15	0.14	0.20
	Log A	0.17	0.17	0.23

Landuse had a MI value of 0.27, indicating a moderate relationship with i_c. This indicates that even knowledge of landuse provides more information about infiltration parameters compared to soil texture, but still less than WRB reference soil groups. Landuse influences infiltration parameters through surface cover, vegetation and human activities such as tillage or construction, which can alter soil structure and permeability. This moderate MI value reflects the significant, but not exclusive, role of landuse in determining infiltration rates.

An almost similar trend was observed when comparing classifiers with 3D infiltration parameters. The MI value obtained between texture and K_s was 0.15, indicating low explanatory power for the variability of texture-based classification of K_s (Table 6). This aligns with the evaluation of soil texture impacts on 1D infiltration data, where also i_c-texture showed the lowest MI value. In similar manner, the MI was low for the land-use based classification of infiltration data (Table 6) where K_s −landuse pairs received a MI value of 0.14. This essentially means that soil texture and landuse are not representative enough for the infiltration parameters and especially for i_c and K_s. Similarly, the MI values were low for the texture and land-use classifications of A and S parameters (Table 6). In contrast, and in line with 1D infiltration parameters, WRB showed higher correlation with all 3D infiltration parameters (K_s, S, and A), meaning WRB reference soil groups were more informative about infiltration processes than soil texture and landuse. With MI values of 0.20 for K_s, 0.26 for S and 0.23 for A, WRB has approximately 33 % higher MI value than soil texture (with MI value of 0.16 for all 3D infiltration parameters). Yet, and even with an MI lower than that of WRB, soil texture still contributes moderately to variability of infiltration parameters but is limited in its ability to encompass the chemical and structural complexity experienced within soils, a feature that WRB captures readily due to its hierarchical nature. On the other hand, landuse, although relevant, is frequently an indirect contributor. It influences infiltration by changing surface conditions or organic matter but does not explicitly consider inherent soil properties.

When comparing the MI obtained between the WRB reference groups and different infiltration parameters, overall 3D parameters obtained lower MI values compared to 1D parameters, implying a higher variability of these 3D infiltration parameters. This is probably due to the higher complexity of 3D water flow in the soil with varying contributions of lateral flow compared to 1D flow, which is limited to vertical downward flow only. As discussed earlier, another source of variability

that may add to the complexity of 3D infiltration parameters is likely to be the variety of instruments/infiltrometers used to characterize 3D infiltration data, each with different set-ups and requirements. In contrast, 1D infiltration data is usually only obtained by double ring and is less subjective compared to other instruments/infiltrometers.

It is worth noting that although MI analysis quantifies the coupling between infiltration parameters and different soil classifiers, it does not yet account for physical soil processes like structure, aggregation or biopores. While these are widely recognized as features with a strong control on infiltration behavior (e.g., Vereecken et al., 2016), they are not contained in the current SWIG database. As is noted for many other large-scale soil datasets, our ability to mechanistically interpret the patterns we observed becomes limited. We therefore emphasize the importance of building structural attributes into future database developments, to enable more process-based analyses and to better tie statistical associations to underlying soil hydrological processes.

3.4. Integrated classification through combination of texture, landuse, and WRB classes

Our results reveal that, although WRB shows higher representativeness when examining infiltration parameter variability within different classes, we hypothesize that representativeness can still be further improved by combining information on soil texture, landuse, and WRB classes. Hence, we used a combined classification to test again the variability of infiltration parameters (both 1D and 3D) among different classes. In addition, we also developed a simple model using a decision tree algorithm where i_c (in the case of 1D) and K_s (in the case of 3D) were predicted using classical inputs of commonly used pedo-transfer functions (e.g., clay, silt, sand, OC, and Db) with and without inclusion of landuse and WRB information.

Our results indicate a clear improvement in the representativeness of infiltration parameters when landuse and soil texture classes are combined with the WRB classification (Fig. 4 and Tables S1 and S2). For i_c, the MI value for WRB alone was 0.51, while the MI value for the combined classification (landuse + soil texture + WRB) was significantly higher at 0.66. This shows an increase of 0.15 in the MI value when landuse and soil texture classes are combined with WRB. Similarly, for K_s, the MI value for WRB alone was 0.21, and the MI value for the combined parameters was 0.54. This represents a substantial increase of 0.33 in the MI value when landuse and soil texture classes are added to the WRB classification. These comparisons demonstrate that while WRB shows a higher individual representativeness for both 1D and 3D infiltration parameters compared to landuse and soil texture, the combination of all three classification systems provides a considerably more informative representation. The increase in MI values for the combined parameters highlights the complementary nature of these classifications and their collective ability to better explain the variability in soil infiltration characteristics.

3.5. Role of WRB and landuse classification in infiltration modeling

The results of our simple modeling exercise (Table 7, Figs. 5–7) show that excluding landuse and WRB from the feature set leads the RF model to achieve R^2 values of 0.773, 0.517, and 0.817 for K_s, S, and A in the training subset, and 0.373, 0.339, and 0.364 for K_s, S, and A in the test data. The RMSE values K_s, S, and A were 2.0, 1.82, and 1.81 for the training set and 3.42, 1.81, and 3.67 for the test set. This means that the model fitted the training data quite well but had difficulty generalizing to the unseen data in the test set. The low R^2 in the test dataset combined with a higher RMSE indicates that the model could not capture how complex the infiltration process is without important environmental variables. The difference in training and test performance can also be attributed to overfitting, which in turn is likely due to the weak explanatory power of the selected features.

In addition to the above modeling configuration, we developed additional RF models (one for each of K_s, S, and A parameters) by incorporating information on landuse and WRB classifications, which

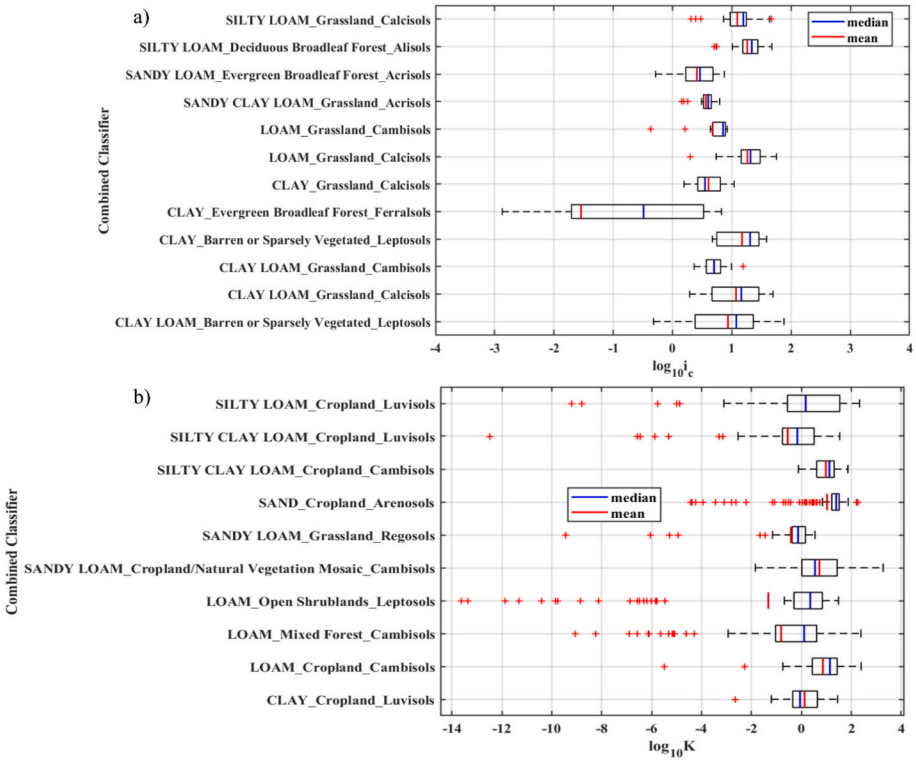


Fig. 4. Boxplots showing the distribution of soil infiltration parameters (log-scale) for 1D and 3D infiltration flow across different classes of integrated classification. a) final infiltration rate (*log10* i_c) in *cm/h* and b) saturated hydraulic conductivity (K_s) in *cm/h*. The blue and red lines represent the median and mean values for each combined class. Outliers are indicated by red plus signs.

Table 7
Accuracy of the rando forest models trained and tested with and without WRB reference groups and landuse classes as categorical predictors.

Parameters¥	Without WRB and landuse R^2	RMSE	With WRB and landuse R^2	RMSE	Improvement€ [%] R^2 [test subset]
log *Ks*	0.773 (0.373)	2.00 (3.42)	0.704 (0.455)	2.37 (3.31)	22.9^{ns}
log *S*	0.517 (0.339)	1.82 (1.81)	0.604 (0.368)	1.68 (1.98)	3.2^{ns}
log *A*	0.817 (0.364)	1.81 (3.67)	0.726 (0.496)	2.31 (3.23)	36.3^{ns}

¥: values outside of the parentheses are for training subset and those inside the parentheses are for the evaluation subset.
€: This column shows the percentage improvement in the R^2 value of the test subset when the WRB and landuse classes are used as categorical predictors versus when they are not.
ns: non-significant.

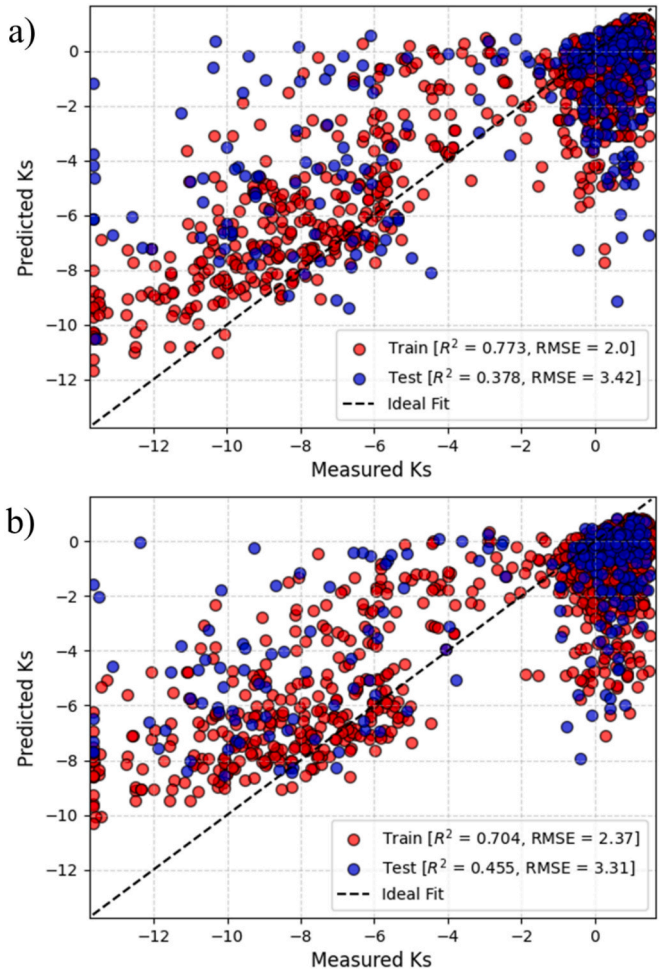


Fig. 5. Measured K_s (in log scale) vs those predicted by Random Forest model. a) represents the predictions using classical inputs of commonly used pedo-transfer functions (e.g., clay, silt, sand, organic carbon, and bulk density) and b) represents the predictions using landuse and WRB soil classification as additional predictors. Training data in red and test data in blue. Dashed line is a 1:1 relationship between the two variables.

significantly increased the performance of the model. While the R^2 for the training data was almost the same as before, it increased considerably for the test datasets for nearly for all models, with a higher R^2 for K_s and *A* (with R^2 values of 0.455 and 0.496) and to some extent also for S (with an R^2 value of 0.368) compared to the previous RF model. This

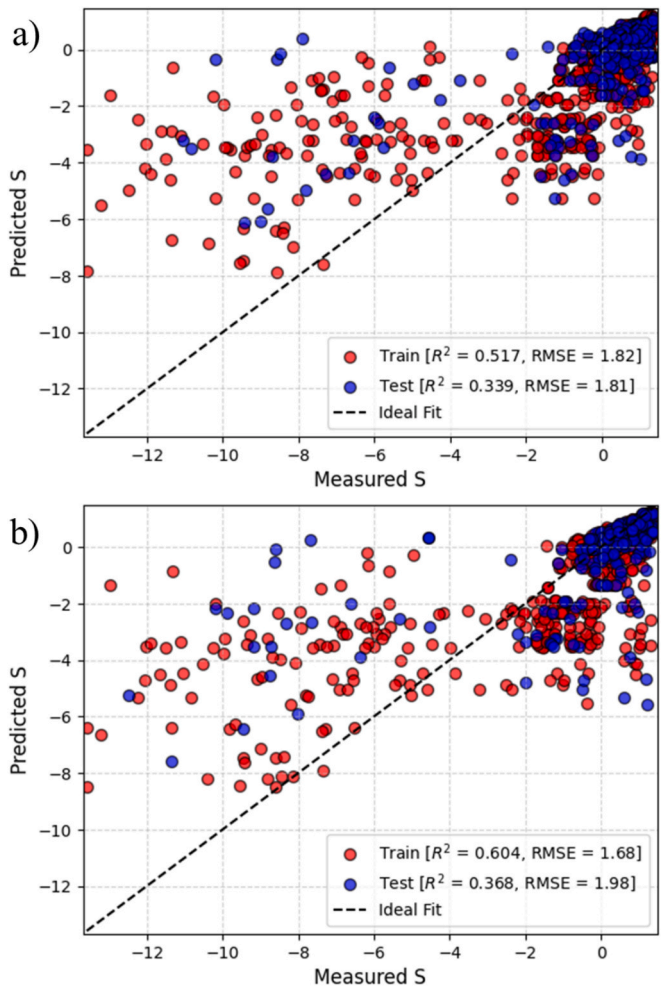


Fig. 6. Measured *S* (in log scale) vs those predicted by Random Forest model. a) represents the predictions using classical inputs of commonly used pedo-transfer functions (e.g., clay, silt, sand, organic carbon, and bulk density) and b) represents the predictions using landuse and WRB soil classification as additional predictors. Training data in red and test data in blue. Dashed line is a 1:1 relationship between the two variables.

simply means a 23 % and 36 % improvement in model performance when facing unseen test data, as well as a minor 3 % improvement in the S parameter. These improvements in a relatively simple RF model underline the need to consider landuse and WRB as important explanatory variables that can explain a substantial part of the spatial and physical variability of infiltration characteristics.

When comparing the models, the addition of landuse and WRB clearly results in a more informative feature space that allows the model to identify more accurate relationships between inputs and infiltration parameters. This is consistent with previous studies that have emphasized the role of land cover and soil classification systems in influencing hydrological processes (Groenendyk et al., 2015). Despite the improvements, it should be noted that the model still shows signs of overfitting, as shown by the large discrepancy between training and test metrics in the second exercise. This could be due to the high sensitivity of RF models to complex or categorical variables with high cardinality or possibly to an imbalance in the data. As a remedial strategy, further work could investigate the use of cross-validation, hyperparameter tuning or alternative modeling approaches such as gradient boosting or neural networks.

The absolute increase in model performance from adding WRB reference groups and landuse classes was still a substantive improvement over the baseline models — especially in *A* and K_s on previously

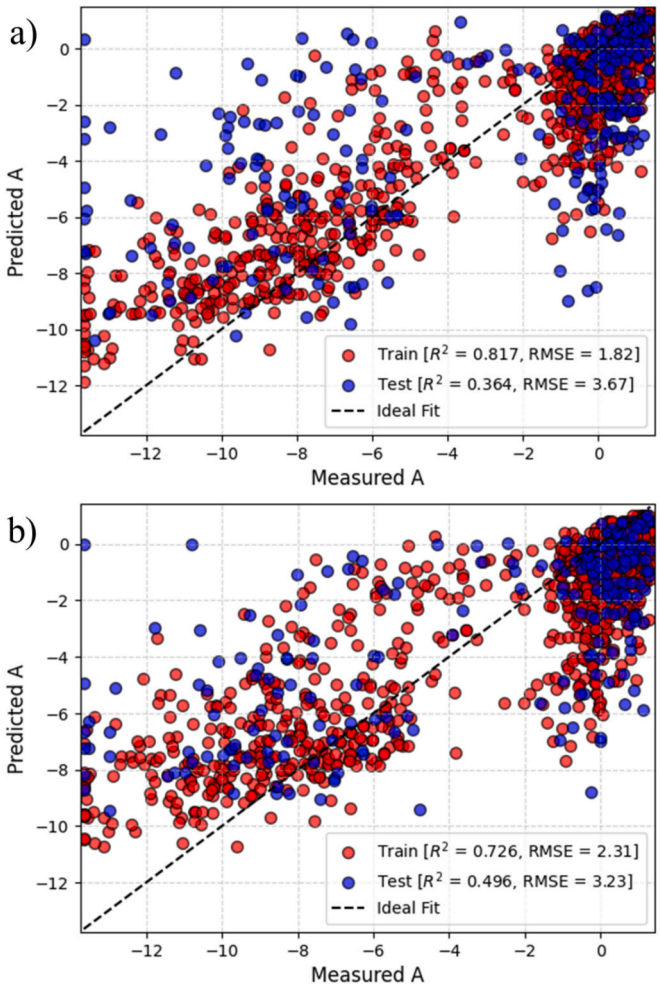


Fig. 7. Measured *A* (in log scale) vs those predicted by Random Forest model. a) represents the predictions using classical inputs of commonly used pedotransfer functions (e.g., clay, silt, sand, organic carbon, and bulk density) and b) represents the predictions using landuse and WRB soil classification as additional predictors. Training data in red and test data in blue. Dashed line is a 1:1 relationship between the two variables.

unseen test data. These are important gains, because of three factors:

- **The simplicity of the modeling setup** — We only tried simple RF models, few predictors, no really advanced feature engineering nor deep hyperparameter tuning. This also suggests that the signal coming from that type of data are indeed useful and did not just derive from endless hyperparameter fiddling.
- **The SWIG dataset is also relatively large and extremely heterogeneous**: therefore, signal-to-noise ratios are inherently low. Given the noisy nature of these datasets, any gains (modest as they may be) likely point to powerful underlying relations that can be leveraged in more controlled or regional contexts.
- **Proof-of-concept for process-informed PTFs**: Our results confirm the premise of the MI analysis. PTFs bounded to categorization systems that capture environmental history and management context can explain variability in infiltration beyond what textural or bulk-density data alone can.

The small improvement in S with either method may reflect this parameter, being mainly sensitive to short-term surface conditions (e.g., initial soil moisture, crusting, vegetation cover) that are not reproduced by WRB or landuse categories. This would further underline the importance of surface-sensitive covariates in future data collection campaigns to complement the stable classification of information.

From an applied perspective, there can be considerable value in even modest improvements in predictive accuracy for K_s and A as small changes in infiltration parameterization may have large impacts on runoff predictions, water balance estimates and erosion modeling over long time scales or large areas when utilized within hydrological and land surface models. As such, although the gains here may appear not to be transformative by naive statistical standards, they represent a strategic gain: the successful demonstration that soil and landuse can then be useful in large-scale PTF development where we have often been too timid in using them as categorically. It suggests that these factors should be systematically integrated in future modeling exercises along with more complex land cover characteristics of the soil surface, and it provides a conceptual link to a more extensive model analysis we intend to develop in subsequent work.

4. CONCLUSION AND OUTLOOK

This study analyses the variability of common infiltration parameters within and between soil textures, landuse types, and WRB reference groups, using a global infiltration database of SWIG. Soil texture is a common classification system for infiltration parameters (i_c and K_s), yet our results show that there is substantial within-class variability, indicating it may not be as effective as a primary predictor of infiltration parameters. The influence of landuse on infiltration parameters is comparable to that of soil texture in terms of the level of explained variability, but high variability within landuse classes diminishes its predictive utility. Variability in soil infiltration parameters, however, could be better explained by using WRB reference groups, which have less within-class variability and greater mutual information with infiltration parameters.

Our analysis emphasizes that the triple challenges of landuse, soil texture, and a defined classification of soils reference groups need to be considered through an integrated classification approach to enhance the representation of infiltration parameters in ecological and hydrological models. This was also proven by including the WRB and landuse types as predictors in the developed model using the Random Forest algorithm, which resulted in a higher model accuracy compared to the same model without the inclusion of the WRB and landuse types as predictors. The results could offer crucial guidance for designing efficient and effective frameworks for land and water resource management. Despite their relatively minor absolute impact on model performance, these results demonstrate that the categorical soil and landuse information is valuable for proper estimation of infiltration parameters. Although modest in relative terms, the predictive gains represent a marginal improvement of potential application to large spatially/temporally heterogeneous datasets like SWIG where infiltration estimates substantially impact predictions for runoff generation, water balance and erosion. In doing so, not only was a stable system of classification integrated, but a series of dynamic descriptors directly related to land surface characteristics were successfully implemented to provide the requisite data input for calibration/evaluation of PTFs, which ultimately led to improved prediction performance and in turn demonstrate the need for both stable classification systems and dynamic abiotic properties in PTF development if robustness across environmental gradients is sought.

Future research needs to figure out how taxonomic soil structure information can be linked to measured soil structural properties. Additional research also should consider both temporal variability and changing climate conditions to improve predictive models. In this regard, division of the dataset according to climate-oriented latitude classes — Equatorial (0°S–10°N), Tropical (10°–23.5° N/S), Subtropical (23.5°–35° N/S), Temperate (35°–50° N/S), Boreal/Cool-temperate (50°–66.5° N/S), and Polar (>66.5° N/S) — revealed well-defined differences in the spread of 1D infiltration rates but less-significant differences in the case of 3D infiltration parameters. Although further extensive stratification of the data points may be necessary for future

research, the calculation of MI for the aforementioned climate-oriented latitude classes confirmed that they are much more informative for 1D infiltration parameters predictions (with a MI value of 0.1–0.3) and less informative for 3D infiltration parameters (with an MI value of less than 0.07). Despite the discouraging results for the 3D infiltration parameters, which may be due to more complex flow behavior and hidden contributing factors, these findings highlight the potential benefits of using climate-specific stratification for large-scale infiltration assessments.

While our findings clearly show that soil groups contain substantial information to explain the variation observed in infiltration parameters, the underlying mechanisms driving these differences remain to be fully understood. Many of the diagnostic features used to define soil groups, particularly in systems such as the WRB, focus on subsurface characteristics that may not directly influence surface infiltration processes. Given that our infiltration experiments were limited to the uppermost 10 cm of the soil profile, it is possible that the properties influencing infiltration at this depth are not fully captured by conventional soil classifications. The WRB highlights that top-soil characteristics are dynamic and prone to rapid change, and therefore often receive limited emphasis in classification schemes. However, previous efforts have proposed complementary systems for topsoil classification (Broll et al., 2006; Jabiol et al., 2013; Zanella et al., 2018), which could provide more direct insights into surface-related hydrological behavior. While we do not explore this aspect further in the present study, identifying specific topsoil attributes that govern infiltration across different soil groups represents a promising direction for future research.

Apart from the surface perspective adopted in this study, we anticipate that future studies of soil profiles will consider vertical variation more noticeably. Heterogeneous and horizon-specific hydraulic properties may have a significant impact on infiltration behavior compared to the homogeneous topsoil assumption. Although our dataset and analysis were restricted to infiltration in the top 10 cm of soil, improvements in non-invasive geophysical methods (e.g. surface and downhole surveys) offer promising opportunities for characterizing vertical stratification at relatively low cost. These datasets could be used to incorporate this information into infiltration modelling through corresponding parameterizations (e.g. harmonic or geometric averages) or depth-dependent hydraulic functions, thereby enhancing both empirical and physically based PTF approaches. Integrating this vertical component would represent a significant step forward in bridging the gap between large-scale statistical analyses and mechanistic, profile-resolved flow models.

It should be noted that although our results show that soil texture, landuse, and WRB reference groups explain some of the observed variability in infiltration parameters, important physical factors such as soil structure, macroporosity and biogenic channels are likely responsible for much of the variance left unexplained. Yet directly affecting preferential flow pathways and infiltration dynamics, these features are missing from the SWIG database and many other large-scale datasets. Thus, our interpretations also are of a statistical nature rather than mechanistic. We thus conclude with a call for future database development and measurement campaigns that directly consider these properties to help introduce structural/biological controls into upscaling of infiltration databases. Integration of this information could not only enhance the association to statistical benchmarks and underlying process but also assist parameterization in land surface and hydrological models.

We also recognize that initial soil moisture conditions are critical for controlling infiltration behavior, especially K_s. However, approximately 70 % of SWIG records lack this information, leaving us with a low number of points available for statistical analysis within each classifier category. Therefore, we were unable to stratify or begin with different initial moisture levels without compromising the study's generalizability. This limitation may explain some of the variability in our results that we could not account for, suggesting that initial moisture levels should be recorded as part of future infiltration data. This would lead to better parameterization of infiltration processes in land surface and hydrological models and allow for a more accurate interpretation of observed patterns.

Although our study primarily focused on the challenges of upscaling when estimating infiltration parameters over larger spatial extents, we recognize that pore-scale processes, such as macropores, biogenic channels, and fractures, influence the dynamics of infiltrating water. These processes strongly regulate preferential flow and infiltration rates, yet they are not directly represented in the SWIG database. Therefore, bridging the gap between pore-scale modelling (e.g., Blunt, 2001; 2017) and large empirical datasets is a key area of future research. In this context, we used the term 'capillarity' in a way that is universally applicable to sorptivity-driven infiltration. However, this should be interpreted as being applicable mostly in the presence of an adequate surface water supply. For drier conditions, gravity and Darcian resistance dominate capillary forces, and film flow may be more pertinent than classical capillarity. Future research may benefit from integrating field-scale infiltration observations back to the pore scale ('downscaling') wherever possible, both to improve the accuracy with which multiscale controlling factors of infiltration dynamics can be predicted and to enhance the mechanistic basis of upscaling.

In conclusion, our study stresses the need for Supplemental information to soil texture for estimating infiltration parameters. Immediate implications for practitioners and researchers are (1) to account for WRB reference groups and landuse in model applications of infiltration, (2) to record initial soil moisture and structural attributes as part of field measurements more systematically, and (3) apply other classification systems in concert to improve the identification surface controls on infiltration. Applying these measures may increase both the precision of hydrological models and the effectiveness of land and water resources planning strategies.

CRediT authorship contribution statement

S. Farnaz Sharghi: Writing – review & editing. **Sara L. Bauke:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Mehdi Rahmati:** Writing – review & editing, Methodology, Data curation. **Dymphie J. Burger:** Writing – review & editing. **Harry Vereecken:** Writing – review & editing. **Wulf Amelung:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the German Science Foundation (Deutsche Forschungsgemeinschaft DFG) – SFB 1502/1-2022 – project number 450058266, as well as by core funds from the Soil Science and Soil Ecology group, University of Bonn.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2025.117550.

Data availability

the data used in this research is already publicly avaliable at https://doi.org/10.1594/PANGAEA.885492

References

Alaoui, A., Caduff, U., Gerke, H.H., Weingartner, R., 2011. Preferential flow effects on infiltration and runoff in grassland and forest soils. Vadose Zone J. 10 (1), 367–377.

Angulo-Jaramillo, R., Vandervaere, J.-P., Roulier, S., Thony, J.-L., Gaudet, J.-P., Vauclin, M., 2000. Field measurement of soil surface hydraulic properties by disc and ring infiltrometers: a review and recent developments. Soil Tillage Res. 55 (1–2), 1–29.

Athmann, M., Kautz, T., Banfield, C., Bauke, S., Hoang, D.T., Lüsebrink, M., Pausch, J., Amelung, W., Kuzyakov, Y., Köpke, U., 2017. Six months of L. terrestris L. activity in root-formed biopores increases nutrient availability, microbial biomass and enzyme activity. Appl. Soil Ecol. 120, 135–142.

Autovino, D., Bagarello, V., Iovino, M., Lassabatere, L., Yilmaz, D., 2024. Parameterizing Haverkamp Model from the Steady-State of Numerically Generated Infiltration: Influence of Algorithms for Steady-State selection. Hydrol. Process. 38 (11), e15330.

Basset, C., Abou Najm, M., Ghezzehei, T., Hao, X., Daccache, A., 2023. How does soil structure affect water infiltration? a meta-data systematic review. Soil Tillage Res. 226, 105577.

Ben-Hur, M., Yolcu, G., Uysal, H., Lado, M., Paz, A., 2009. Soil structure changes: aggregate size and soil texture effects on hydraulic conductivity under different saline and sodic conditions. Soil Res. 47 (7), 688–696.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. J. Hydrol. 249 (1–4), 11–29.

Blunt, M.J., 2001. Flow in porous media—pore-network models and multiphase flow. Curr. Opin. Colloid Interface Sci. 6 (3), 197–207.

Blunt, M.J., 2017. Multiphase flow in permeable media: a pore-scale perspective. Cambridge University Press.

Bogner, C., Gaul, D., Kolb, A., Schmiedinger, I., Huwe, B., 2010. Investigating flow mechanisms in a forest soil by mixed-effects modelling. Eur. J. Soil Sci. 61 (6), 1079–1090.

Bouma, J., 2006. Hydropedology as a powerful tool for environmental policy research. Geoderma 131 (3–4), 275–286.

Broll, G., Brauckmann, H.J., Overesch, M., Junge, B., Erber, C., Milbert, G., Baize, D., Nachtergaele, F., 2006. Topsoil characterization—recommendations for revision and expansion of the FAO-Draft (1998) with emphasis on humus forms and biological features. J. Plant Nutr. Soil Sci. 169 (3), 453–461.

Carsel, R.F., Parrish, R.S., 1988. Developing joint probability distributions of soil water retention characteristics. Water Resour. Res. 24 (5), 755–769.

Cheng, Q., Tang, C.-S., Xu, D., Zeng, H., Shi, B., 2021. Water infiltration in a cracked soil considering effect of drying-wetting cycles. J. Hydrol. 593, 125640.

Ciglasch, H., Amelung, W., Totrakool, S., Kaupenjohann, M., 2005. Water flow patterns and pesticide fluxes in an upland soil in northern Thailand. European Journal of Soil Science 56 (6), 765–777.

Clapp, R.B., Hornberger, G.M., 1978. Empirical equations for some soil hydraulic properties. Water Resour. Res. 14 (4), 601–604.

Clausnitzer, V., Hopmans, J., Starr, J., 1998. Parameter uncertainty analysis of common infiltration models. Soil Sci. Soc. Am. J. 62 (6), 1477–1487.

Cosby, B., Hornberger, G., Clapp, R., Ginn, T., 1984. A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. Water Resour. Res. 20 (6), 682–690.

Costa, A.D., Albuquerque, J.A., Costa, A.D., Pértile, P., Silva, F.R.D., 2013. Water retention and availability in soils of the State of Santa Catarina-Brazil: effect of textural classes, soil classes and lithology. Rev. Bras. Ciên. Solo 37, 1535–1548.

de Souza, S.E., de Almeida, I.K., Carvalho, G.A., Alves Sobrinho, T., Vianna Bacchi, C.G., 2019. Influence of soil properties and environmental characteristics in water infiltration in urban areas. Geociências 38 (4), 1029–1038.

Dos Santos, K.F., Barbosa, F.T., Bertol, I., de Souza Werner, R., Wolschick, N.H., Mota, J. M., 2018. Study of soil physical properties and water infiltration rates in different types of land use. Sem. Ciên. Agrár. 39 (1), 87–97.

Edwards, W., Shipitalo, M., Norton, L., 1988. Contribution of macroporosity to infiltration into a continuous corn no-tilled watershed: implications for contaminant movement. J. Contam. Hydrol. 3 (2–4), 193–205.

Fashi, F.H., Gorji, M., Shorafa, M., Sarmadian, F., Mohammadi, M.H., 2010. Evaluation of some infiltration models and hydraulic parameters. Spanish Journal of Agricultural Research (1),, 210–217.

Fu, B., Chen, L., Ma, K., Zhou, H., Wang, J., 2000. The relationships between land use and soil conditions in the hilly area of the loess plateau in northern Shaanxi, China. Catena 39 (1), 69–78.

Gan, T., 2024. Jupyter Notebooks for the soilgrids Data Component.

Ghorbani Dashtaki, S., Homaee, M., Mahdian, M.H., Kouchakzadeh, M., 2009. Site-dependence performance of infiltration models. Water Resour. Manag. 23, 2777–2790.

Groenendyk, D.G., Ferre, T.P., Thorp, K.R., Rice, A.K., 2015. Hydrologic-process-based soil texture classifications for improved visualization of landscape function. PLoS One 10 (6), e0131299.

Haghnazari, F., Shahgholi, H., Feizi, M., 2015. Factors affecting the infiltration of agricultural soils. Int. J. Agron. Agric. Res. 6 (5), 21–35.

Hao, H.-X., Wei, Y.-J., Cao, D.-N., Guo, Z.-L., Shi, Z.-H., 2020. Vegetation restoration and fine roots promote soil infiltrability in heavy-textured soils. Soil Tillage Res. 198, 104542.

Haverkamp, R., Ross, P., Smettem, K., Parlange, J., 1994. Three-dimensional analysis of infiltration from the disc infiltrometer: 2. Physically based infiltration equation. Water Resour. Res. 30 (11), 2931–2935.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One 12 (2), e0169748.

Horton, R.E., 1939. Analysis of runoff-plat experiments with varying infiltration-capacity. EOS Trans. Am. Geophys. Union 20 (4), 693–711.

Horton, R.E., 1941. An Approach Toward a Physical Interpretation of Infiltration-Capacity. Soil Sci. Soc. Am. J. 5 (C), 399–417.

Jabiol, B., Zanella, A., Ponge, J.-F., Sartori, G., Englisch, M., Van Delft, B., De Waal, R., Le Bayon, R.-C., 2013. A proposal for including humus forms in the World Reference Base for Soil Resources (WRB-FAO). Geoderma 192, 286–294.

Kargas, G., Londra, P.A., Sotirakoglou, K., 2021. Saturated hydraulic conductivity measurements in a loam soil covered by native vegetation: Spatial and temporal variability in the upper soil layer. Geosciences 11 (2), 105.

Kögel-Knabner, I., Amelung, W., 2021. Soil organic matter in major pedogenic soil groups. Geoderma 384, 114785.

Lassabatere, L., Peyneau, P.-E., Yilmaz, D., Pollacco, J., Fernández-Gálvez, J., Latorre, B., Moret-Fernández, D., Di Prima, S., Rahmati, M., Stewart, R.D., 2021. Mixed formulation for an easy and robust numerical computation of sorptivity. Hydrol. Earth Syst. Sci. Discuss. 2021, 1–23.

Latorre, B., Moret-Fernández, D., Lassabatere, L., Rahmati, M., López, M., Angulo-Jaramillo, R., Sorando, R., Comín, F., Jiménez, J., 2018. Influence of the β parameter of the Haverkamp model on the transient soil water infiltration curve. J. Hydrol. 564, 222–229.

Lei, T., Liu, H., Pan, Y., Zhao, J., Zhao, S., Yang, Y., 2006. Run off-on-out method and models for soil infiltrability on hill-slope under rainfall conditions. Sci. China Ser. D 49, 193–201.

Libohova, Z., Schoeneberger, P., Bowling, L.C., Owens, P.R., Wysocki, D., Wills, S., Williams, C.O., Seybold, C., 2018. Soil systems for upscaling saturated hydraulic conductivity for hydrological modeling in the critical zone. Vadose Zone J. 17 (1), 1–20.

Liu, B., Xie, G., Zhang, X., Zhao, Y., Yin, X., Cheng, C., 2015. Vegetation root system, soil erosion and ecohydrology system: A review, 2015 International Forum on Energy, Environment Science and Materials. Atlantis Press, pp. 271-279.

Liu, Y., Evans, J., McCabe, M.F., de Jeu, R.A., Van Dijk, A., Su, H., 2010. Influence of cracking clays on satellite estimated and model simulated soil moisture. Hydrol. Earth Syst. Sci. 14 (6), 979–990.

Ma, D., Shao, M., 2008. Simulating infiltration into stony soils with a dual-porosity model. Eur. J. Soil Sci. 59 (5), 950–959.

Ma, W., Zhang, X., Zhen, Q., Zhang, Y., 2016. Effect of soil texture on water infiltration in semiarid reclaimed land. Water Qual. Res. J. Canada 51 (1), 33–41.

Ma, X., Zhang, B., Shi, D., Lu, G., 2007. Study on Soil Infiltration Characteristic of Different Land Utilization Types in Purple Soil Hilly Region. Journal of Soil and Water Conservation (5),, 25–29.

Mantel, S., Dondeyne, S., Deckers, S., 2023. World reference base for soil resources (WRB). Goss, Margaret Oliver Encyclopedia of Soils in the Environment, 2nd ed.; Michael, J., Ed, 206-217.

Mao, L., Li, Y., Hao, W., Mei, X., Bralts, V.F., Li, H., Guo, R., Lei, T., 2016. An approximate point source method for soil infiltration process measurement. Geoderma 264, 10–16.

Marquart, A., Eldridge, D.J., Geissler, K., Lobas, C., Blaum, N., 2020. Interconnected effects of shrubs, invertebrate-derived macropores and soil texture on water infiltration in a semi-arid savanna rangeland. Land Degrad. Dev. 31 (16), 2307–2318.

Mirzaee, S., Zolfaghari, A.A., Gorji, M., Dyck, M., Ghorbani Dashtaki, S., 2014. Evaluation of infiltration models with different numbers of fitting parameters in different soil texture classes. Arch. Agron. Soil Sci. 60 (5), 681–693.

Moret-Fernández, D., Latorre, B., López, M., Pueyo, Y., Lassabatere, L., Angulo-Jaramilo, R., Rahmati, M., Tormo, J., Nicolau, J., 2020. Three-and four-term approximate expansions of the Haverkamp formulation to estimate soil hydraulic properties from disc infiltrometer measurements. Hydrol. Process.

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. Trans. ASABE 58 (6), 1763–1785.

Moyroud, N., Portet, F., 2018. Introduction to QGIS. QGIS Generic Tools 1, 1–17.

Muñoz, D.R., Villas, D.B., Arnedo, M.T.E., Negrell, M.A.G., Monreal, N.-L.-R., León, J., Romero, M.E.N., Pardini, G., Muela, M.P.S., 2017. Analysing the effect of land use and vegetation cover on soil infiltration in three contrasting environments in northeast Spain. Cuadernos De Investigación Geográfica: Geograph. Res. Lett. 43, 141–169.

Mwendera, E., Saleem, M.M., 1997. Infiltration rates, surface runoff, and soil loss as influenced by grazing pressure in the Ethiopian highlands. Soil Use Manag. 13 (1), 29–35.

Naidu, R., Rengasamy, P., 1993. Ion interactions and constraints to plant nutrition in Australian sodic soils. Soil Res. 31 (6), 801–819.

Návar, J., Mendez, J., Bryan, R.B., Kuhn, N.J., 2002. The contribution of shrinkage cracks to bypass flow during simulated and natural rainfall experiments in northeastern Mexico. Can. J. Soil Sci. 82 (1), 65–74.

Or, D., Keller, T., Schlesinger, W.H., 2021. Natural and managed soil structure: on the fragile scaffolding for soil functioning. Soil Tillage Res. 208, 104912.

Pachepsky, Y., Karahan, G., 2022. On shapes of cumulative infiltration curves. Geoderma 412, 115715.

Pachepsky, Y., Park, Y., 2015. Saturated hydraulic conductivity of US soils grouped according to textural class and bulk density. Soil Sci. Soc. Am. J. 79 (4), 1094–1100.

Parlange, J.-Y., Lisle, I., Braddock, R., Smith, R., 1982. The three-parameter infiltration equation. Soil Sci. 133 (6), 337–341.

Rahmati, M., 2017. Reliable and accurate point-based prediction of cumulative infiltration using soil readily available characteristics: a comparison between GMDH, ANN, and MLR. J. Hydrol. 551, 81–91.
Rahmati, M., Latorre, B., Lassabatere, L., Angulo-Jaramillo, R., Moret-Fernández, D., 2019. The relevance of Philip theory to Haverkamp quasi-exact implicit analytical formulation and its uses to predict soil hydraulic properties. J. Hydrol. 570, 816–826.
Rahmati, M., Latorre, B., Moret-Fernández, D., Lassabatere, L., Talebian, N., Miller, D., Morbidelli, R., Iovino, M., Bagarello, V., Neyshabouri, M.R., 2022. On infiltration and infiltration characteristic times. Water Resour. Res. 58 (5), e2021WR031600.
Rahmati, M., Vanderborght, J., Šimůnek, J., Vrugt;, J.A., Moret-Fernández, D., Latorre, B., Lassabatere, L., Vereecken, H., 2020. Soil hydraulic properties estimation from one-dimensional infiltration experiments using characteristic time concept. Vadose Zone J. 19 (1), e20068.
Rahmati, M., Weihermüller, L., Vanderborght, J., Pachepsky, Y.A., Mao, L., Sadeghi, S. H., Moosavi, N., Kheirfam, H., Montzka, C., Van Looy, K., 2018a. Development and analysis of the Soil Water Infiltration Global database. Earth Syst. Sci. Data 10 (3), 1237–1263.
Rahmati, M., Weihermüller, L., Vereecken, H., 2018b. Soil Water Infiltration Global (SWIG) Database, Supplement to: Rahmati, M et al.(2018): Development and Analysis of Soil Water Infiltration Global Database. Earth System Science Data, PANGAEA, https://doi. org/10.1594/PANGAEA 885492.
Reichenberger, S., Amelung, W., Laabs, V., Pinto, A., Totsche, K.U., Zech, W., 2002. Pesticide displacement along preferential flow pathways in a Brazilian Oxisol. Geoderma 110 (1–2), 63–86.
Rice, R., Milczarek, M., Keller, J., 2014. A Critical Review of Single Ring Cylinder Infiltrometers with Lateral Flow Compensation.
Ross, B.C., 2014. Mutual information between discrete and continuous data sets. PLoS One 9 (2), e87357.
Salehi, M.H., 2018. Challenges of Soil Taxonomy and WRB in classifying soils: some examples from iranian soils. Bull. Geogr. Phys. Geogr. Ser. 14, 63–70.
Schaap, M.G., Leij, F.J., 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci. 163 (10), 765–779.
Schad, P., 2016. The international soil classification system WRB, 2014. Novel methods for monitoring and managing land and water resources in Siberia, 563-571.
Schaefer, C.E.G., Fabris, J.D., Ker, J., 2008. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. Clay Miner. 43 (1), 137–154.
Schneider, F., Don, A., 2019. Root-restricting layers in German agricultural soils. Part II: Adaptation and melioration strategies. Plant and Soil 442, 419–432.
Shoji, S., Fujiwara, Y., Yamada, I., Saigusa, M., 1982. Chemistry and clay mineralogy of Ando soils, Brown forest soils, and Podzolic soils formed from recent Towada ashes, northeastern Japan. Soil Sci. 133 (2), 69–86.
Šimůnek, J., van Genuchten, M.T., Šejna, M., 2008. Development and applications of the HYDRUS and STANMOD software packages and related codes. Vadose Zone J. 7 (2), 587–600.
Šimůnek, J., Van Genuchten, M.T., Šejna, M., 2016. Recent developments and applications of the HYDRUS computer software packages. Vadose Zone J. 15 (7).
Smettem, K., 1987. Characterization of water entry into a soil with a contrasting textural class: spatial variability of infiltration parameters and influence of macroporosity. Soil Sci. 144 (3), 167–174.
Smettem, K., Parlange, J., Ross, P., Haverkamp, R., 1994. Three-dimensional analysis of infiltration from the disc infiltrometer: 1. a capillary-based theory. Water Resour. Res. 30 (11), 2925–2929.
Soil Survey Staff, 2022. Keys to Soil Taxonomy. USDA Natural Resources Conservation Service.
Stürmer, S.L.K., Dalmolin, R.S.D., Azevedo, A.C.d., Pedron, F.d.A., Menezes, F.P., 2009. Relação da granulometria do solo e morfologia do saprolito com a infiltração de água em Neossolos Regolíticos do rebordo do Planalto do Rio Grande do Sul. Ciência Rural 39, 2057–2064.
Suárez, E., Arcos, E., Moreno, C., Encalada, A., Álvarez, M., 2013. Influence of vegetation types and ground cover on soil water infiltration capacity in a high-altitude páramo ecosystem. Av. Cienc. Ing. 5, B14–B21.
Sun, D., Yang, H., Guan, D., Yang, M., Wu, J., Yuan, F., Jin, C., Wang, A., Zhang, Y., 2018. The effects of land use change on soil infiltration capacity in China: a meta-analysis. Sci. Total Environ. 626, 1394–1401.
Tejedor, M., Neris, J., Jiménez, C., 2013. Soil properties controlling infiltration in volcanic soils (Tenerife, Spain). Soil Sci. Soc. Am. J. 77 (1), 202–212.
Todisco, F., Vergni, L., Iovino, M., Bagarello, V., 2023. Changes in soil hydrodynamic parameters during intermittent rainfall following tillage. Catena 226, 107066.
Törmänen, T., Lindroos, A.-J., Ilvesniemi, H., Starr, M., 2020. Development of Podzols in relation to Jenny's soil formation factors. EGU Gener. Assem. Conf. Abstr. 21440.
Van de Genachte, G., Mallants, D., Ramos, J., Deckers, J., Feyen, J., 1996. Estimating infiltration parameters from basic soil properties. Hydrol. Process. 10 (5), 687–701.
Van Es, H., Cassel, D., Daniels, R., 1991. Infiltration variability and correlations with surface soil properties for an eroded Hapludult. Soil Sci. Soc. Am. J. 55 (2), 486–492.
Vereecken, H., Amelung, W., Bauke, S.L., Bogena, H., Brüggemann, N., Montzka, C., Vanderborght, J., Bechtold, M., Blöschl, G., Carminati, A., 2022. Soil hydrology in the Earth system. Nat. Rev. Earth Environ. 3 (9), 573–587.
Vereecken, H., Schnepf, A., Hopmans, J.W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M.H., Amelung, W., Aitkenhead, M., 2016. Modeling soil processes: Review, key challenges, and new perspectives. Vadose Zone J. 15 (5), vzj2015.
Vereecken, H., Weihermüller, L., Assouline, S., Šimůnek, J., Verhoef, A., Herbst, M., Archer, N., Mohanty, B., Montzka, C., Vanderborght, J., 2019. Infiltration from the pedon to global grid scales: an overview and outlook for land surface modeling. Vadose Zone J. 18 (1), 1–53.
Vrugt, J.A., Hopmans, J.W., Gao, Y., Rahmati, M., Vanderborght, J., Vereecken, H., 2024. The time validity of Philip's two-term infiltration equation: an elusive theoretical quantity? Vadose Zone J., e20309
Wang, L., Zhong, C., Gao, P., Xi, W., Zhang, S., 2015. Soil infiltration characteristics in agroforestry systems and their relationships with the temporal distribution of rainfall on the loess Plateau in China. PLoS One 10 (4), e0124767.
Wardak, D.L.R., Padia, F.N., De Heer, M.I., Sturrock, C.J., Mooney, S.J., 2022. Zero tillage has important consequences for soil pore architecture and hydraulic transport: a review. Geoderma 422, 115927.
Weihermüller, L., Lehmann, P., Herbst, M., Rahmati, M., Verhoef, A., Or, D., Jacques, D., Vereecken, H., 2021. Choice of Pedotransfer Functions matters when Simulating Soil Water Balance Fluxes. J. Adv. Model. Earth Syst. 13 (3), e2020MS002404.
Woignier, T., Morell, M., Primera, J., Duffours, L., 2008. Correlation between large water content and fractal structure in volcanic soils. Proc. BALWOIS 1–10.
Yilmaz, D., Lassabatere, L., Moret-Fernández, D., Rahmati, M., Angulo-Jaramillo, R., Latorre, B., 2023. Soil-dependent β and γ shape parameters of the Haverkamp infiltration model for 3D infiltration flow. Hydrol. Process. 37 (6), e14928.
Yu, M., Zhang, L., Xu, X., Feger, K.H., Wang, Y., Liu, W., Schwärzel, K., 2015. Impact of land-use changes on soil hydraulic properties of Calcaric Regosols on the Loess Plateau, NW China. J. Plant Nutr. Soil Sci. 178 (3), 486–498.
Zanella, A., Ponge, J.-F., Jabiol, B., Sartori, G., Kolb, E., Gobat, J.-M., Le Bayon, R.-C., Aubert, M., De Waal, R., Van Delft, B., 2018. Humusica 1, article 4: Terrestrial humus systems and forms—Specific terms and diagnostic horizons. Appl. Soil Ecol. 122, 56–74.
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.
Zimmermann, B., Elsenbeer, H., De Moraes, J.M., 2006. The influence of land-use changes on soil hydraulic properties: Implications for runoff generation. For. Ecol. Manage. 222 (1–3), 29–38.