ORIGINAL RESEARCH ARTICLE

Vadose Zone Journal

Analyzing dual porosity in soil hydraulic properties using soil databases for pedotransfer function development

Yonggen Zhang[1,3] | **Lutz Weihermüller**[2,3] | **Brigitta Toth**[4] | **Muhammad Noman**[2] | **Harry Vereecken**[2,3]

[1]Institute of Surface-Earth System Science, School of Earth System Science, Tianjin Univ., Tianjin 300072, China

[2]Agrosphere Institute IBG 3, Forschungszentrum Jülich GmbH, Jülich 52425, Germany

[3]International Soil Modelling Consortium (ISMC)

[4]Centre for Agricultural Research, Institute for Soil Sciences, Herman Ottó út 15, Budapest 1022, Hungary

Correspondence
Yonggen Zhang, Institute of Surface-Earth System Science, School of Earth System Science, Tianjin Univ., Tianjin, 300072, China.
Email: ygzhang@tju.edu.cn

Assigned to Associate Editor Hailong He.

Abstract

Current databases of soil hydraulic properties (SHPs) have typically been used to develop pedotransfer functions (PTFs) to estimate water retention [$\theta(h)$] assuming a unimodal pore-size distribution. However, natural soils often show the presence of bimodal to multimodal pore-size distributions. Here, we used three widely spread databases for PTF development: UNsaturated SOil hydraulic DAtabase (UNSODA) 2.0, Vereecken, and European hydropedological data inventory (EU-HYDI), to analyze the presence of structural effects in both $\theta(h)$ and hydraulic conductivity [$K(h)$]. Only undisturbed samples were included in the analysis that contained enough datapoints for both $\theta(h)$ and $K(h)$ properties, especially in the wet range. One-hundred ninety-two samples were suitable for our analysis, which is only 1% of the total samples in the three databases. Results showed that 65% of the samples exhibited a bimodal pore-size distribution, and bimodality was not limited to fine-textured but also coarser-textured soils. The Mualem–van Genuchten (MvG) expression for both unimodal and bimodal soils was not able to predict the observed unsaturated K. Only a joint fitting of measured $\theta(h)$ and $K(h)$ functions provided parameter estimates that were able to describe unsaturated K for uni- and bimodal soils. In addition, we observed a negative relationship between α and n in the case of low sand content (<52%) for both unimodal and bimodal matrix domain properties, contradicting the classical notion. The ratio of α for the macropore and matrix domain was positively correlated with the fraction of macropores and sand content. We anticipate that the results will contribute to deriving PTF for structured soils and avoid unrealistic combinations of MvG parameters.

1 | INTRODUCTION

Soil hydraulic properties (SHPs) are key in understanding and predicting local to continental-scale hydrological, energy-related, and biogeochemical processes occurring at the land surface in managing ecosystem services of soils and for the optimal and sustainable design of agricultural systems. Yet, SHPs are only sparsely available and their measurement is tedious and time-consuming. Pedotransfer functions (PTFs) are used to estimate these soil hydraulic properties, that is, the water retention [$\theta(h)$] characteristic and hydraulic conductivity [$K(h)$] functions from simple soil properties like soil particle size distribution (sand, silt, and clay content), soil organic carbon content, and bulk density (e.g., Clapp & Hornberger, 1978; Van Looy et al., 2017; Vereecken et al., 2010; Wösten et al., 2001). These functions avoid the need for direct measurement of SHPs and facilitate the application of mathematical models that describe water, energy, and carbon fluxes in soils at various spatial scales ranging from the field to the global scale. Currently, the estimation of the soil $\theta(h)$ characteristic and the $K(h)$ function using PTFs is based on the assumption that soils exhibit a unimodal pore size distribution. There is, however, ample experimental evidence in the literature that soils may show a hierarchical organization of soil structure leading to bi- or multimodal pore size distributions (Brewer, 1964; Hadas, 1987; Dexter, 1988; Dexter et al., 2008; Oades & Waters, 1991).

Early experimental studies [e.g., Germann and Beven (1981) and Othmer et al. (1991)] analyzed the effect of macropore structures on soil hydrological processes. In the work of Germann and Beven (1981), the authors used large undisturbed soil monoliths to study the effect of soil macroporosity on the infiltration capacity of soils. The main finding was that the $K(h)$ decreased by factors of 4.3 and 18 when macropore flow was excluded, pointing to the importance of macropores for water translocation at and close to saturation. Another experimental approach was used by Othmer et al. (1991), who analyzed the presence of a bimodal pore size distribution in the soil $\theta(h)$ characteristic measured on undisturbed soil samples. Here, they concluded that the calculated $K(h)$ functions derived from the bimodal retention characteristic agreed better with the measured hydraulic conductivities in the field than those obtained by neglecting the bimodal distribution in the retention characteristic.

Concepts to describe water flow in bi-and multimodal soils were proposed by Addiscott (1977), Beven and Germann (1981), Jarvis et al. (1991), and Gerke and van Genuchten (1993), amongst others. In the work of Addiscott (1977), a conceptual model accounting for the effect of soil structure on water flow and leaching of solutes was developed, whereby he divided the flow of the soil solution into a mobile and a retained phase, implicitly recognizing the presence of two pore-size domains. Beven and Germann (1981) proposed

Core Ideas

- Only 1% of the samples in the three databases was suitable for analyzing the presence of bimodality.
- Bimodality was seen in 65% of the samples, both in fine-textured and coarser-textured soils.
- Joint fitting of $\theta(h)$ and $K(h)$ data outperforms the fitting of only $\theta(h)$ and using MvG function to estimate $K(h)$.
- Contradicting the classical notion, a negative relationship between the shape parameters α and n was observed for both uni- and bimodal properties.
- α for the macropore and matrix domains was positively correlated with the fraction of macropores and sand content.

a one-dimensional model of bulk flow in soils exhibiting a micro- and macropore domain using the Richards equation. Jarvis et al. (1991) also developed a model for water and solute transport for macroporous soils, whereby the model considers again a micro- and a macropore domain. In contrast to those of Beven and Germann (1981), the water flow in the macropore domain is solved based on Darcy's law with saturated hydraulic conductivity (K_s) expressed as a function of microporosity in a simple power-law form, whereas Richards equation is solved for the matrix domain using Brooks and Corey (1964) and Mualem (1976) for the parameterization of $\theta(h)$ and unsaturated $K(h)$ characteristic, respectively. To describe the exchange between both domains, an empirical interaction term was added. Gerke and van Genuchten (1993) proposed a dual-porosity model to account for the bimodal nature of pore size distributions in natural soils by considering a macropore or fracture pore domain and a less permeable matrix pore domain, whereby the Richards equation is solved for both domains. A first-order rate equation was additionally introduced to transfer solutes between the macropore and matrix pore domains. For the parameterization of the soil hydraulic characteristics, the Mualem–van Genuchten model (Mualem, 1976; van Genuchten, 1980) was used to parametrize water flow in both domains.

Another avenue to account for bi- or multimodal soils is integrating this process directly into the $\theta(h)$ and $K(h)$ function. A first attempt to quantify the effect of macropores on water flow was conducted by Philip (1968) using analytical solutions of a dual-porosity model, but suitable concepts using modified retention and conductivity functions accounting for macropore flow or bi- or multimodal soils were developed much later by Smettem and Kirkby (1990), Ross and Smettem (1993), Durner (1994), and Dexter and Richard (2009).

Smettem and Kirkby ([1990]) used the Mualem–van Genuchten model to describe the water flow in the intra-aggregate (macropore) and in the interaggregate (micropore) domain of a structured soil and the matching points (θ_m, h_m) and (K_m, h_m) at the boundary between both domains, where θ_m is the soil water content at the boundary, h_m the matric potential at the matching point, and K_m the unsaturated $K(h)$ at the matching point. Each domain was characterized by its specific van Genuchten shape parameters α (relating to the inverse of air entry pressure) and n (measuring the pore-size distribution). Ross and Smettem ([1993]), on the other hand, introduced a bimodal parameterization of the $\theta(h)$ and $K(h)$ function by superimposing individual retention and conductivity curves (here the authors denote this as overlapping pore-size distributions), whereby one curve accounts for the macropore domain and the other one for the matrix domain. The authors also stated that the use of sums of simple functions has the advantages of simplicity and predicting power and that the effects of macroporosity can be represented in a physically realistic manner. The same concept was also taken up by Durner ([1994]) for the Mualem–van Genuchten characteristics and extended by Priesack and Durner ([2006]) for the closed-form equation between $\theta(h)$ and $K(h)$ function, whereby this parameterization has been widely used since then (e.g., Spohrer et al., [2006]; Jadoon et al., [2012]; Schwen et al., [2014]). Finally, Dexter and Richard ([2009]) published a model for $\theta(h)$ in bimodal soils, which was further developed into a model for trimodality accounting for different levels or types of pores. This model is based on the exponential (Boltzmann) $\theta(h)$ function and can be extended to include any number of modes of porosity as may exist because of the hierarchical nature of soil structure. Compared with the Durner model, this model has not been widely used yet.

Despite the fact that the pore network of structured soils may often be best described by bi- or multimodal pore-size distributions and that bi- or multimodal soil hydraulic functions may better describe flow in structured soils, no effort has been undertaken to develop PTFs that estimate bi- or multimodal soil hydraulic properties. This is caused by the lack of suitable variables that are able to predict hydraulic parameters describing the soil macropore domain and by the double number of parameters for bimodal model fitting requires more measurement points of $\theta(h)$ characteristic or the $K(h)$ function than provided in most data collections. Additionally, measurement points in the wet range of the $K(h)$ are often lacking. But even if data of both characteristics [$\theta(h)$ and $K(h)$] were available, most PTFs were developed on retention data only (Weihermüller et al., [2021]). Until today, almost all PTFs use a unimodal description of the soil $\theta(h)$ characteristic to estimate soil hydraulic parameters in analytical expressions of the $\theta(h)$ relationship and to derive the unsaturated hydraulic conductivity from the soil $\theta(h)$

characteristic using, for example, Mualem–van Genuchten or Brooks and Corey equations. Some exceptions are Li et al. ([2014]) and Haghverdi et al. ([2020]), who estimated the parameters of bimodal soil $\theta(h)$ models from basic soil properties. Both works call attention to the limited application of the derived PTFs because of the low number of training samples used for the analysis. In using unimodal approaches to fit soil hydraulic properties of structured soils, Dexter and Richard ([2009]) notice that "it seems likely to us that the need for adjustable 'shape factors', for example the parameters m and n that are used in the van Genuchten ([1980]) $\theta(h)$ equation, is mainly a consequence of trying to fit unimodal models to data from soils that are inherently bi- or multimodal as a consequence of the hierarchical nature of soil structure (Brewer, [1964]; Hadas, [1987]; Dexter, [1988]; Oades & Waters, [1991])."

Despite the limited number of soil samples to analyze the bimodal and multimodal properties and limited efforts devoted to developing PTFs for these properties, it is generally accepted that hydraulic properties of fine-textured soils are strongly influenced by structural properties such as aggregates and macropores (Coppola, [2000]; Coppola, Basili, et al., [2009]; Coppola, Comenga, et al. [2009]). Multiporosity and dual-porosity/dual-permeability models have been proposed to better account for these properties. For example, Coppola ([2000]) superimposed a van Genuchten or a Ross and Smettem formulation to describe the macroporosity and van Genuchten for the matrix part. Through validation on 18 clay–clay loam soils, they were found to be more closely correlated with measured soil $\theta(h)$ and unsaturated conductivity dataset when bimodal approaches were used compared with unimodal approaches. Coppola, Basili, et al. ([2009]) investigated the effects of unimodal and bimodal soil hydraulic properties on the predictions of soil water content in an ensemble framework. They found significantly better simulation of soil water content compared with measurement by using a bimodal approach when validated on clayey loam soils. Most of the work regarding bimodal soil hydraulic properties has mostly been tested and validated on fine-textured soils. Here we will analyze the structural properties from a wide range of soil textures and check if the effect of structural properties on soil hydraulic properties is only limited to the fine-textured soils or occurs across all textural classes.

Lehmann et al. ([2020]) suggested that traditional estimation of soil hydraulic properties by PTFs is unconstrained and may result in unphysical soil hydraulic parameters, for example, the unphysical relationship between α and n parameters in the van Genuchten model. By injecting an additional soil-specific characteristic length that controls capillary flow to an evaporating surface, the occurrence of unphysical soil hydraulic parameter combinations was reduced, and the land surface representation was improved. It is, therefore, interesting to check if the α and n relationship is physically reasonable by analyzing the estimated soil hydraulic parameters obtained

from the unimodal and bimodal soil hydraulic functions fitted to soil hydraulic characteristic data.

The aims of this study are (a) to collect and establish a database of soil hydraulic properties for the analysis of the effect of soil structure on soil $\theta(h)$ characteristics and $K(h)$, (b) to evaluate unimodal and bimodal soil hydraulic functions using the Mualem–van Genuchten model with respect to the estimated parameters, and (c) to evaluate the dependence of the van Genuchten shape factors α and n with respect to different fitting strategies.

2 | MATERIALS AND METHODS

2.1 | Soil hydraulic functions

Two different soil functions, namely the unimodal and bimodal model, were used in this study to describe the soil hydraulic properties by parameter estimation from measured retention and saturated and unsaturated hydraulic conductivities. For the first approach, the commonly applied Mualem–van Genuchten (MvG) functions (Mualem, 1976; van Genuchten, 1980) were fitted to the data, whereby the effective volumetric saturation (S_e) is described by the following equation:

$$S_e(h) = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \begin{cases} \frac{1}{[1+|\alpha h|^n]^m} & h < 0 \\ 1 & h \geq 0 \end{cases} \quad (1)$$

where θ (cm^3 cm^{-3}) is the volumetric water content; h (cm) is the matric potential; α (cm^{-1}), n, and m are empirical shape parameters, where m is set to $m = 1 - 1/n$ in this study as suggested by van Genuchten (1980); and θ_s and θ_r are the volumetric water contents as the saturated and residual water contents (cm^3 cm^{-3}), respectively. During the parameter estimation described in the Objective Functions and Parameter Estimation section, a large majority of soils yielded a θ_r value of 0. To be consistent with all fittings, θ_r was set to 0 for all soil samples, or in other words, θ_r was not fitted at all, as was also done by, for example, Weynants et al. (2009).

The $K(h)$ function is expressed by Equation 2:

$$K(h) = K_s S_e^\tau \left[1 - \left(1 - S_e^{1/m}\right)^m \right]^2 \quad (2)$$

where K_s is the $K(h)$ (cm d^{-1}) acting as a matching point, whereby classically, the saturated $K(h)$ is used for it (Schaap & Leij, 2000), and τ is the tortuosity factor.

The second approach used to describe the soil hydraulic properties is the bimodal model according to Durner (1994), which accounts for soil structure and macroporosity especially for soils at high water contents. In this model, the porous soil can be divided into i overlapping MvG functions

weighted by a factor w_i. The effective saturation can therefore be described by Equation 3:

$$S_e(h) = \sum_{i=1}^k w_i \left(\frac{1}{1+|\alpha_i h|^{n_i}} \right)^{m_i} \quad (3)$$

with the sum of w_1 to w_k being equal to 1. The bimodal model is therefore restricted by $k = 2$. Combining Equation 3 with the pore size distribution model of Mualem (1976), the bimodal unsaturated $K(h)$ function can be expressed according to Priesack and Durner (2006) by the following equation:

$$K(S_e) = K_s \left(\sum_{i=1}^k w_i S_{ei} \right)^\tau \left\{ \frac{\sum_{i=1}^k w_i \alpha_i \left[1 - \left(1 - S_e^{1/m_i}\right)^{m_i} \right]}{\sum_{i=1}^k w_i \alpha_i} \right\}^2 \quad (4)$$

Both the soil hydraulic parameters and the weighting factors w_i were determined in the parameter estimation and are described in Objective Functions and Parameter Estimation section.

2.2 | Soil $\theta(h)$ and $K(h)$ database

Three different databases were used, namely the Vereecken (Vereecken et al., 1989, 1990), the UNSODA 2.0 database (Nemes et al., 2001), and the EU-HYDI database as described by Weynants et al. (2013). The UNSODA 2.0 database included some samples from Vereecken et al. (1989, 1990) and was further extended by other datasets covering soil samples from the global distribution.

All databases were heterogeneous and consisted of data from different sources and measurement procedures and essential soil variables such as soil texture (sand, silt, and clay content), bulk density, soil organic carbon content if included in the dataset. Water retention and $K(h)$ data were available at various levels of detail. The Vereecken and UNSODA 2.0 databases initially have 183 and 790 soil samples, respectively, whereas the EU-HYDI has 18,537 soil samples. Only data from undisturbed samples were used in our study, reducing the number of samples to 183, 520, and 14,294, respectively. For the Vereecken dataset, the water content in the low-pressure head range was determined for the undisturbed samples, while water content in the high-pressure head range (>246 kPa) was measured on disturbed samples taking from core samples. For the UNSODA and EU-HYDI dataset, no clear measurement method is available for the pressure ranges, and we analyzed the soil samples that was labeled as 'undisturbed'. In addition, $\theta(h)$ and unsaturated $K(h)$ data had to be available, and we analyzed the soil samples measured in the laboratory, reducing the number of soil samples to 145, 244, and 1,412, respectively, for the three

databases. To ensure that enough information is included in the measured data, at least three $\theta(h)$ data pairs at matric potential higher than -20 cm are required including θ_s. For $K(h)$, three pairs at matric potential higher than -20 cm are required, whereby one should be K_s. Scotter (1978) suggested that soil pores with water-entry pressure higher than -15 cm can be classified as macropore, whereas Jarvis (2007) suggested the value as -10 to -6 cm. We relaxed the requirement of data pairs near the saturation to -20 cm and two more data pairs in addition to θ_s and K_s to make sure that sufficient information is included near saturation. Additionally, the samples should have at least six $\theta(h)$ pairs and at least seven $K(h)$ pairs in the entire matric potential range as partially proposed by Schaap and van Genuchten (2006). The number of data pairs will guarantee that the number of observations will be larger than the number of fitted parameters, avoiding underdetermined cases and a nonunique solution of the fitted parameters. By setting these constraints, the Vereecken database consists of 59, the UNSODA database of 16, and the EU-HYDI of 119 individual samples totaling 194 samples, accounting for only 1% of the total number samples in the three databases.

2.3 | Objective functions and parameter estimation

For the fitting of the unimodal and bimodal soil $\theta(h)$ equations (Equation 1 or 3) and $K(h)$ functions (Equation 2 or 4), we minimize the objective function of the sum of squared residuals (SSR) between estimated and observed quantities of water content and $K(h)$ respectively at given matric potential.

For the joint fitting of the unimodal MvG equations (joint of Equations 1 and 2) and bimodal equations (joint of Equations 3 and 4), the objective function is the sum of the weighted SSR for $\theta(h)$ and $K(h)$ as follows:

$$\Phi(\mathbf{P}) = \sum_{i=1}^{I} u_i \left[\theta_i(h_i) - \theta_i'(h_i) \right]^2$$
$$+ \sum_{j=1}^{J} v_j \left[\log K_j(h_j) - \log K_j'(h_j) \right]^2 \quad (5)$$

where $\theta_i(h_i)$ and $K_j(h_j)$ are respectively the ith measured water content and jth $K(h)$ data for each soil sample; $\theta_i'(h_i)$ and $K_j'(h_j)$ are corresponding estimated water content and hydraulic conductivity; I and J are the number of measured $\theta(h)$ and $K(h)$ for each sample; $\mathbf{P}$ is the MvG parameter vector for unimodal and MvG parameter vector and weighting factors w_i for bimodal equations, being used to estimate $\theta_i'(h_i)$ and $K_j'(h_j)$ at corresponding h; and u_i and v_j are weighting terms used to balance the soil $\theta(h)$ and $K(h)$, which are respectively obtained from the inverse of the variance of the unimodal (or bimodal for bimodal fitting) soil $\theta(h)$ and $K(h)$

functions fitted to the data, that is, the aforementioned SSR values. Large SSR values are weighted less compared with the small SSR values. See Weynants et al. (2009) and Zhang and Schaap (2017) for the details of deriving the balance terms.

A combination of optimization algorithms was used to minimize the objective function for each soil sample. To obtain the optimal (global minimum) soil hydraulic parameters, a step-wise fitting approach was performed where the dataset was firstly fitted by using the generalized simulated annealing algorithm in R package (Xiang et al., 2013), the particle swarm optimization described by Kennedy and Eberhart (1995) and built-in PSO package (Bendtsen, 2012), and the differential evolution approach described by Mullen et al. (2011) and Price et al. (2006) implemented in DEoptim package (Ardia et al., 2020). The corresponding fitting with the lowest SSR values among the three algorithms was selected and used as an initial guess to estimate parameters for the second fitting. During the second fitting, the particle swarm optimization algorithm was used because this algorithm was found to perform somewhat superior over the other algorithms and was computationally efficient. The second fitting results were used as the final optimum fitted parameters. These procedures were repeated for both the $\theta(h)$ and $K(h)$ datasets to obtain the $\theta(h)$ and $K(h)$ parameters. The joint fitting of dual-porosity soil $\theta(h)$ curves and dual-porosity soil $K(h)$ curves was fitted to the dataset again using the same second-fitting approach. Here, it has to be noted that the second-fitting might lead to minor improvement compared with the first fitting, and it is worth deciding on the need for a second fitting for specific samples. The SSR values from the aforementioned dual-porosity soil $\theta(h)$ model and dual-porosity $K(h)$ model were used to weight the joint fitting and balance the $\theta(h)$ and $K(h)$ dataset. The obtained dual-porosity soil hydraulic parameters were used as the final parameters.

During the fitting process, a large majority of soils yielded a θ_r value of 0. To be consistent with all fittings, θ_r was set to 0 for all soil samples, or in other words θ_r was not fitted at all, as was also done by Weynants et al. (2009), for example. In addition, because of the limited number of measurement of soil characteristic data, less-fitted parameters and the simpler representation of soil hydraulic properties are more desirable. For the joint fitting of retention and $K(h)$ data, the tortuosity (τ) was also considered as a fitting parameter as was done by Schuh and Cline (1990) or Schaap and Leij (2000), for example.

During the fitting, the parameter bounds used were set to $0.00001 < \alpha < 1$ (cm^{-1}), $1.01 < n < 30.0$, $0.001 < \theta_s < 1.0$ (cm^3 cm^{-3}), $0.0001 < K_s < 10,000,000$ (cm d^{-1}), and $0.0 < w < 1.0$, whereby w is the weighting factor for the macropore (fracture) domain (See Equation 3). For the constraint of τ, Peters et al. (2011) suggested that the fundamental requirement is to keep the hydraulic functions monotonic, that is, both $\theta(h)$ and $K(h)$ curves should decrease or stay constant as

the matric potential h decreases, whereas a stricter physical requirement is to keep the concave shape of the conductivity function. To keep more flexibility for the fitting, we choose to keep the monotonicity of the coupled MvG $K(h)$ function. For the unimodal (Equation 2) fitting, τ should be larger than $-2/m$ (where m is van Genuchten parameter; see Equation 1) suggested by Peters et al. (2011) and the lower and upper bounds of τ is set as $-10.0 < \tau < 10.0$. For bimodal fitting strategies (Equation 4), τ is suggested to be larger than -2 $\min\{|1/m_k|\}$ ($k = 1, 2$, respectively, for macropore and matrix domain) and τ is also set as $-10.0 < \tau < 10.0$. Additionally, a constrain with respect to α_i and n_i was included to clearly identify α and n values for the macropore (α_1 and n_1) and matrix (α_2 and n_2) dominated domain by the following equation:

$$\text{SSR} = 10\text{SSR if} (\alpha_1 < \alpha_2 \ \& \ n_1 > n_2) \ or \ (\alpha_1 > \alpha_2 \ \& \ n_1 < n_2) \tag{6}$$

which ensured that α_1 has to be larger than α_2 and n_1 larger than n_2.

Further to the joint fitting, we also analyzed the accuracy of the $K(h)$ function by considering the parameters resulting from the fitting of the unimodal and bimodal soil $\theta(h)$ model and considering K_s as the measured K_s value (K_s is fixed during the fitting) and τ fixed at 0.5.

All fittings were performed in the R software (Version 3.6.0) (R Core Team, 2020).

2.4 | Dependency of α on n in the van Genuchten equation

To explore the theoretical relationship between α (cm^{-1}) vs. n parameters, the derivative of α vs. n, $\partial\alpha/\partial n$, was calculated for the case where $m = 1 - 1/n$, which leads to the closed form of relative $K(h)$ suggested by van Genuchten (1980). The derivative of α vs. n derived in this section does not show how the two parameters are related when comparing different soils. However, we use this derivative to show how α varies with n in order to match the mathematical relationship between S_e and h.

When αh is significantly >1, Equation 1 can be approximated by the following:

$$S_e \approx (\alpha h)^{1-n} \tag{7}$$

yielding Equation 8:

$$\frac{1}{1 - n} \log S_e = \log(\alpha h) \tag{8}$$

where α is, therefore, given as Equation 9:

$$\alpha = \frac{1}{h} e^{\frac{1}{1-n} \log S_e} \tag{9}$$

thus leading to Equation 10:

$$\frac{\partial\alpha}{\partial n} = \frac{\sqrt[1-n]{S_e} \log S_e}{h(1 - n)^2} \tag{10}$$

The right-hand side of Equation 10 is always < 0, because (a) $0 < S_e < 1$, leading to $\sqrt[1-n]{S_e} > 0$; (b) since $n \neq 1$, we obtain $(1 - n)^2 > 0$; and (c) $\log(S_e) < 0$. It becomes clear that $\frac{\partial\alpha}{\partial n} < 0$. Therefore, α increases with decreasing n.

2.5 | Model selection criterion and comparison of parameters in different models

To select the best model to characterize the measured $\theta(h)$ and $K(h)$ data, the Akaike Information Criterion (AIC) (Akaike, 1974) was used. The AIC estimates the differences between the data's unknown true likelihood function and the fitted likelihood function of the model used. Hereby, the smallest AIC values indicate the best model describing the measured data. The AIC is derived from frequentist probability and was calculated based on the following equation:

$$\text{AIC} = N \ln(\text{SSR}/N) + 2p_m + \frac{2p_m(p_m + 1)}{N - p_m - 1} \tag{11}$$

where N is the number of measured data points (soil $\theta(h)$ and $K(h)$ data pairs), p_m is the number of fitted model parameters, and SSR is the sum of squared residuals. For the joint fitting of the $\theta(h)$ and $K(h)$ functions, the total SSR (the terms shown in Equation 5 without adding the weighting terms) was used in AIC calculation, whereby for the fitting of retention data only, the SSR was also calculated for the $K(h)$ data by the use of the closed-form equation (Equations 2 and 4, respectively, for unimodal and bimodal) and the two individual SSRs were summed up. The best model was chosen as those yielding the lowest AIC value.

For the analysis of differences in fitted soil hydraulic parameters, the comparison of the mean of more groups was performed with the nonparametric Kruskal–Wallis test with adjustment of p values by Benjamini and Hochberg (1995) using R package 'agricolae' (De Mendiburu, 2014) based on the probability $p = .05$.

3 | RESULTS AND DISCUSSION

3.1 | General database statistics and properties

For the selected samples obtained from the UNSODA, Vereecken, and EU-HYDI database, the total number of measurements for $\theta(h)$ and $K(h)$ vary significantly with a

TABLE 1 Number of measurements for θ(h), K(h), and θ(h) + K(h) for the 194 samples extracted from the UNsaturated SOil hydraulic DAtabase (UNSODA), Vereecken, and European hydropedological data inventory (EU-HYDI) databases

	θ(h)	K(h)	θ(h) + K(h)
Mean	19	33	52
Min.	7	7	16
Max.	62	118	148
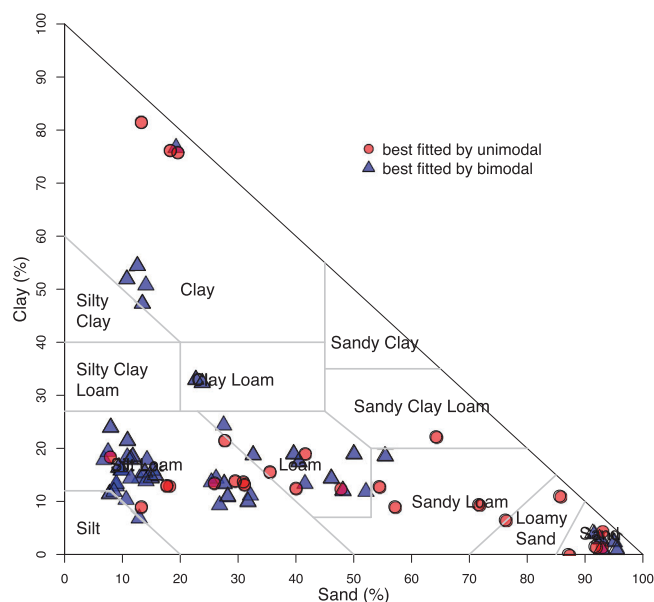
FIGURE 1 USDA soil textural triangle for all 79 soil samples where soil textural information was provided in the database. Red markers indicate those sampled best fitted by unimodal fits, and blue markers indicate bimodal according to the lowest Akaike Information Criterion

minimum of six for θ(h) and seven for K(h), which is slightly larger as the thresholds defined for measurement near saturation [three for θ(h) larger than −20 cm and three for K(h) larger than −20 cm]. The maximum reported data pairs showed an extremely high number of measurements with a number of 62 and 118 individually (see Table 1). Mean data pairs are 19 and 33 for the two characteristics [θ(h) and K(h), respectively].

For the total 194 soil samples, we found that more soil samples were best fitted by the bimodal fit, accounting for 65% of the soil samples. It is interesting and helpful to discover how the best unimodal or bimodal fit varies with different soil textures. To this end, we select soil samples in our database containing soil textural information. Unfortunately, only 79 (41%) samples contain information on soil texture, whereby for those samples, the mean (minimum and maximum) percentages are 33.9% (6.9 and 95.3%) for sand, 47.3% (2.6 and 80.7%) for silty, and 18.8% (0 and 81.4%) for clay, indicating significant variation of soil texture. The soil textural distribution is plotted in Figure 1 for those soils identified

as unimodal (by the lowest AIC values for the unimodal fits, red dot in the figure) and bimodal (by the lowest AIC for the bimodal fits, blue triangle in the figure). It is clearly shown that contrary to the previous observations in the literature that bimodality was often observed in fine-textured soils, heavy clayey soils (clay content >70%) in our dataset did not show a clear bimodality and were mostly best fitted by a unimodal model and median clayey soils (clay content in the range of 40–70%) were best fitted by the bimodal model. In addition, silt loam soil samples, accounting for the largest portion of soil samples containing soil texture information in our database, were best fitted by the bimodal fit. The abundance of silt loam samples that are best fitted with a bimodal model can be explained by the fact that besides silt class, this class has the largest amount of silt compared with all other textural classes, whereby the silt-sized fraction sometimes shows characteristics that are similar to both sand and clay-sized constituents (Lal & Shukla, 2004), which may lead to the development of aggregated soil structures depending on the type of bonding agents present in the soil. These structures are then likely to lead to bi-or multimodal pore size distributions. On the other hand, there is no distinct clustering of unimodal or bimodal soil for the other USDA soil texture classes.

3.2 | Identifying best-fitting strategy

After fitting the data to the unimodal or bimodal model, the best-fitting strategy, that is, fitting to retention data only or the joint fitting of θ(h) and K(h) dataset, was identified. As the classical fitting to estimate the soil hydraulic parameters would be the unimodal fitting to θ(h) data only (abbreviated as unimodal θ(h), hereafter) or jointly fitting of θ(h) and K(h) data with fixed K_s, in which K_s is not fitted but equal to the measured K_s [abbreviated as unimodal θ(h) + K(h)K_s fixed] or K_s fitted [abbreviated as unimodal θ(h) + K(h)K_s fitted], these fitting results are discussed first. The lowest AIC values were still considered as an indicator for the best-fitting strategy.

We found that for the unimodal fitting (i.e., θ(h), θ(h) + K(h) K_s fixed, and θ(h) + K(h) K_s fitted cases) of all soil samples (N = 194), only one sample (sample ID = 101 in the EU-HYDI database; see Supplemental Table S1) was best fitted by using unimodal θ(h) and simultaneously setting τ = 0.5 in predicting the K(h) based on Equation 2. In this case, K_s is fixed during the fitting, and K_s is alternatively called a matching point at saturation in some references (see, e.g., Schaap & van Genuchten, 2006), whereas a measured K_s is often used for the matching point. It is noted that the outperformance of the unimodal θ(h) strategy for this sample is mainly because it has fewer parameters to fit and it has less data points used in the fitting. The unimodal θ(h) strategy has three parameters ($θ_s$, α, n) to fit, whereas unimodal θ(h) + K(h) K_s fixed and θ(h) + K(h) K_s fitted strategies have four ($θ_s$, α, n, τ) and five

TABLE 2 Absolute and relative percentage for the best fitting strategy identified by the lowest Akaike Information Criterion using all data (N = 194)

	Unimodal			Bimodal	
	$\theta(h)$	$\theta(h) + K(h)$ K_s **fixed**	$\theta(h) + K(h)$ K_s **fitted**	$\theta(h)$	$\theta(h) + K(h)$ K_s **fixed**
Total fits	1	123	70	–	–
Percentage	0.5	63.4	36.1	–	–
Total fits	0	48	20	0	126
Percentage	0	24.7	10.3	0	64.9

Note. The first and second rows are the statistics from the unimodal fitting only. The third and fourth rows are the statistics from the combination of the unimodal and bimodal fitting.

(θ_s, α, n, τ, K_s) parameters, respectively. In addition, sample 101 has 12 and 25 data points for the $\theta(h)$ and $\theta(h) + K(h)$ characteristics, respectively. Both fewer parameters to be fitted and less data points available in the fitting for the unimodal $\theta(h)$ strategy lead to the lowest AIC values. In terms of SSR values, the unimodal $\theta(h) + K(h)$ K_s fixed and unimodal $\theta(h) + K(h)$ K_s fitted strategies still outperformed unimodal $\theta(h)$ fitting along with the $K(h)$ prediction. This result is in line with Schaap and Leij (2000), who indicated that this fitting strategy is not suitable to predict the unsaturated hydraulic characteristics, even though this approach is used for many applications because $K(h)$ data is rarely measured. On the other hand, fitting both $\theta(h)$ and $K(h)$ functions at the same time dramatically outperforms the fitting to $\theta(h)$ only, as indicated by lower AIC values for all other samples ($N = 193$) even if fitting both $\theta(h)$ and $K(h)$ simultaneously generated a trade-off between soil $\theta(h)$ and unsaturated $K(h)$ data.

Allowing only unimodal fitting of all soil samples ($N = 194$), the results show that the unimodal $\theta(h) + K(h)$ with K_s fixed strategy yielded the best performance for 63.4% ($N = 123$) of soil samples over all unimodal fitting strategies (Table 2). As in some cases, K_s will be not fixed but fitted; this fitting strategy (unimodal $\theta(h)+K(h)$ K_s fitted) was also performed and yielded the best for 36.1% ($N = 70$) of all samples based on the lowest AIC values. Here, 60.0% ($N = 75$) of the samples fitted best by unimodal $\theta(h) + K(h)$ with K_s fixed and 40.0% ($N = 50$) of the samples fitted best by unimodal $\theta(h) + K(h)$ with K_s fitted are characterized as bimodal. In this case, both options for joint fitting using a unimodal model (unimodal $\theta(h) + K(h)$ with K_s fitted and unimodal $\theta(h) + K(h)$ with K_s fixed) do not yield the overall best fit if bimodal fitting would have also been performed.

If we also allow bimodal fitting and compare all fitting strategies (unimodal and bimodal), the results show that 64.9% ($N = 126$) of the soil samples were best characterized by bimodal, as indicated by the lowest AIC values in Supplemental Table S1 (highlighted in green). In addition, only the fitting based on bimodal $\theta(h) + K(h)$ showed good results in bimodal fitting strategies. The fitting based on bimodal $\theta(h)$ did not show any good result, indicating that most information about bimodality is contained in the conductivity data.

The unimodal $\theta(h) + K(h)$ with K_s fixed strategy showed the best performance for 24.7% ($N = 48$) of soil samples if both unimodal and bimodal strategies were considered (see Table 2). As already stated, in some cases, K_s will also be fitted (not fixed) and this fitting strategy (unimodal $\theta(h) + K(h)$ K_s fitted) showed the best for 10.3% ($N = 20$) of all samples based on the lowest AIC values. An overview of all AIC values for the individual fitting strategies of the 194 soil samples is provided in Supplemental Table S1. The fitting strategy $\theta(h) + K(h)$ K_s fitted for the bimodal fitting was not carried out, as macroporosity information is often included in K_s. Fitting K_s might, therefore, largely cannot accurately characterize the macroporous domain of both the $\theta(h)$ and $K(h)$ function, especially if the data density in the wet range is small. Weynants et al. (2009) has proposed to fit a K_0 [$K(h)$ of matrix flow at zero capillary head], which should be located at a small negative pressure head, for example, −6 cm, to avoid effects of macroporosity on the estimated $\theta(h)$ and $K(h)$ function, as the macroporosity cannot be captured by the classical unimodal MvG equation (e.g., Schaap & van Genuchten, 2006; Weynants et al., 2009). We analyzed the K_0 and compared with the observed K_s in Supplemental Text S1.

Based on these results, it can be concluded that the soils identified as bimodal could have been more accurately described by considering dual-porosity models in the PTF development of Rosetta (Schaap et al., 2001), Rosetta3 (Zhang & Schaap, 2017), in the PTF of Vereecken (Vereecken et al., 1989, 1990) and Weynants (Weynants et al., 2009), as well as those of euptfs (Tóth et al., 2015; Szabó et al., 2021) and Weber et al. (2020), as they used, or partly used, either the UNSODA, Vereecken, or EU-HYDI dataset.

3.3 | Statistical analysis of fitted soil hydraulic parameters

As different fitting strategies will yield different soil hydraulic parameters, we show here the statistics of the estimated parameters α, n, θ_s, τ, and K_s and their relation in different fitting strategies.
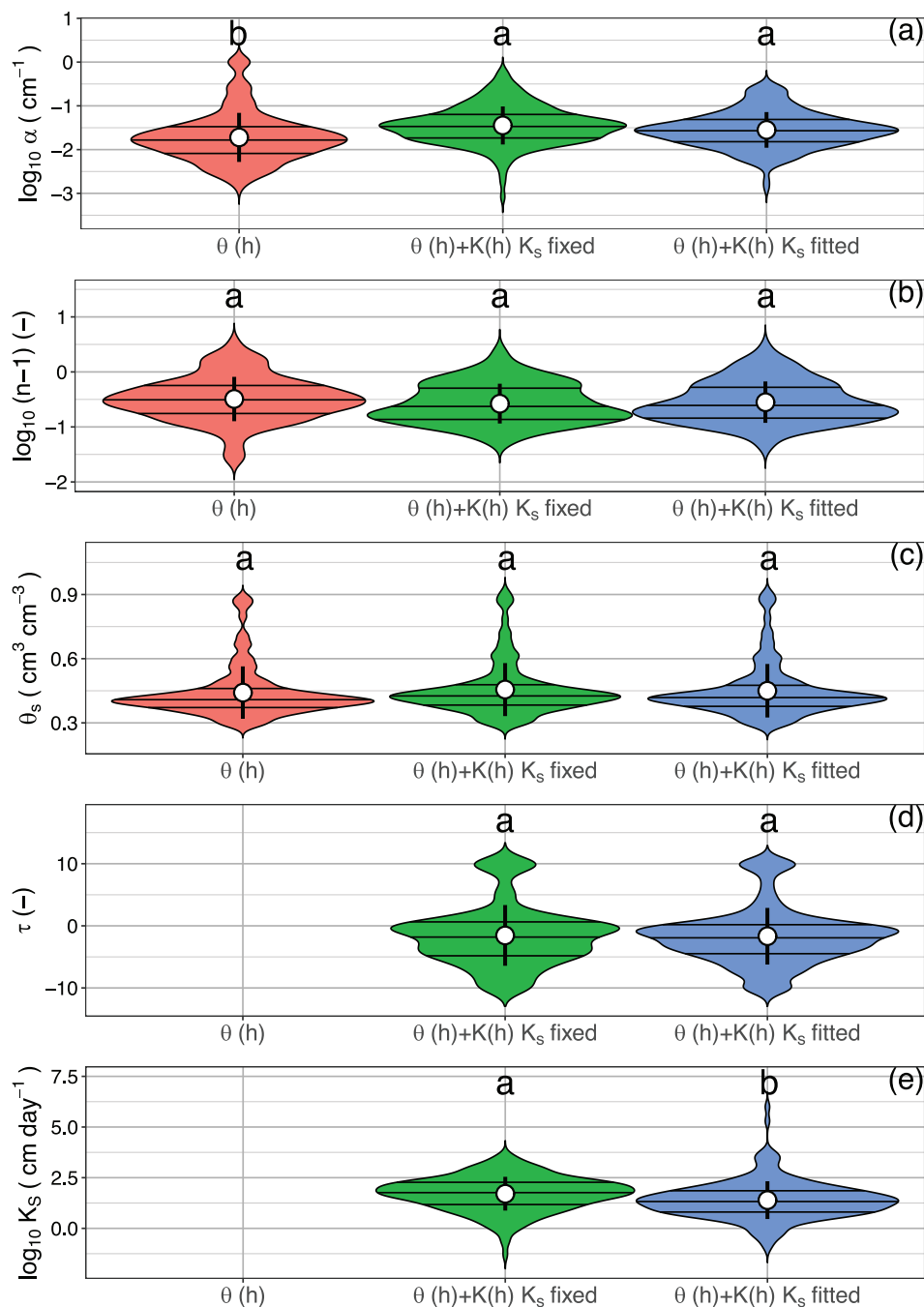
FIGURE 2 Violin plots for Mualem–van Genuchten (a) α, (b) *n*, (c) θ$_s$, (d) τ, and (e) *K*$_s$ in the case of the unimodal fitting strategy using all soil samples (*N* = 194). Centerline indicates the median, lower and upper lines of the 25th and 75th quantiles, the white point the arithmetic mean, and the vertical bars the one standard deviation. Letters above each subfigures indicate significant differences (*p* = .05), whereby cases with different letters are significantly different. It is noted that *K*$_s$ in the θ(*h*) + *K*(*h*) *K*$_s$ fixed case is from the measurement

For the visualization of the statistics of the estimated soil hydraulic parameters, violin plots were generated. In a first step, only the hydraulic parameters of the unimodal fitting for the three unimodal fitting strategies, that is, θ(*h*) only, both hydraulic functions fitted with fixed *K*$_s$ (abbreviated as θ(*h*) + *K*(*h*) *K*$_s$ fixed), and both hydraulic functions fitted with fitted *K*$_s$ (abbreviated as θ(*h*) + *K*(*h*) *K*$_s$ fitted) are depicted in Figure 2 for the fits of all soil samples (*N* = 194). As can

be seen in Figure 2a, the unimodal θ(*h*) fitting strategy yield the smallest median α value (0.017 cm^{-1}, centerline in the figure; see Table 3) compared with the other fitting strategies in which both data sources, θ(*h*) and *K*(*h*), were used (median for θ(*h*) + *K*(*h*) *K*$_s$ fixed = 0.034 cm^{-1} and θ(*h*) + *K*(*h*) *K*$_s$ fitted = 0.027 cm^{-1}). Additionally, the fitted α values from unimodal θ(*h*) are significantly lower (*p* = .05) than in the case of the other two unimodal fitting strategies. On the other hand,

TABLE 3 Fitting statistics of α, n, θ_s, and τ of the unimodal parameters for all soil samples ($N = 194$)

| Parameter | Statistic | Unimodal | | |
		$\theta(h)$	$\theta(h) + K(h)$ K_s fixed	$\theta(h) + K(h)$ K_s fitted
α, cm^{-1}	Arithmetic mean	0.0190	0.0356	0.0282
	Median	0.0169	0.0343	0.0274
	Min.	0.0014	0.0008	0.0013
	Max.	1.0000	0.5680	0.2954
n	Arithmetic mean	1.4944	1.3827	1.4239
	Median	1.3075	1.2214	1.2269
	Min.	1.0244	1.0487	1.0401
	Max.	4.4728	3.6930	4.1332
θ_s, cm^3 cm^{-3}	Arithmetic mean	0.441	0.455	0.450
	Median	0.408	0.426	0.417
	Min.	0.286	0.288	0.289
	Max.	0.887	0.916	0.907
τ	Arithmetic mean	–	−1.542	−1.674
	Median	–	−1.616	−1.760
	Min.	–	−10.000	−10.000
	Max.	–	10.000	10.000

Note. Values for α and n were log$_{10}$ transformed before calculating the mean and median values and then were anti-log transformed for calculating the presented values.

the arithmetic mean of the fitted α value is still the lowest for the fitting to $\theta(h)$ data only (0.019 cm^{-1}) compared with the fitting using both functions (arithmetic mean for $\theta(h) + K(h)$ with K_s fixed = 0.036 cm^{-1} and $\theta(h) + K(h)$ with K_s fitted = 0.028 cm^{-1}). Here, we have to note that we used the anti-log transformed α values for median and mean values to indicate the statistical values mentioned above.

The fitted n value for the different unimodal strategies has median values with 1.494, 1.383, and 1.424, respectively, for $\theta(h)$, $\theta(h) + K(h)$ K_s fixed, and $\theta(h) + K(h)$ K_s fitted, but there is no significant difference between them, which is verified by the same letter above the violin plot in Figure 2b. On the other hand, the mean fitted n values vary with a difference of 0.112 between $\theta(h)$ and $\theta(h) + K(h)$ K_s fixed and 0.071 between $\theta(h)$ and $\theta(h) + K(h)$ K_s fitted (see Table 3). Next, the fitted θ_s was also analyzed (see Figure 2c), whereby the medians are 0.408, 0.426, and 0.417 cm^3 cm^{-3} for unimodal $\theta(h)$, unimodal $\theta(h) + K(h)$ K_s fixed, and unimodal $\theta(h) + K(h)$ K_s fitted, suggesting marginal differences among different fitting strategies.

Finally, differences in the estimated tortuosity factor τ are analyzed, whereby τ was only estimated for the case where $\theta(h)$ and $K(h)$ datasets were simultaneously used for the fitting (see Figure 2d). Factor τ shows considerable variation from the lower limits of −10 to the upper limits of 10. The median values are −1.616 and −1.760 for unimodal $\theta(h) + K(h)$ K_s fixed and unimodal $\theta(h) + K(h)$ K_s fitted strategies, whereas the mean fitted values are −1.542 and −1.674, respectively, for the two strategies (see Table 3). The median and mean val-

ues of τ deviate from the default setting of 0.5, but are close to the optimal value of −1 suggested by Schaap and Leij (2000). The $\theta(h) + K(h)$ K_s fixed strategy produced higher τ values in both the median and mean values. Larger τ yields, in general, higher K values for a given water content or pressure head and also produces a steeper decrease of unsaturated hydraulic conductivity K as the matrix potential h decreases. By setting K_s as fitted in the $\theta(h) + K(h)$ K_s fitted strategy, the optimization procedure results in a more negative τ (see Table 3), and significantly smaller K_s (being the fitted matching point, see Figure 2e for the fitted K_s values) with a less steep unsaturated $K(h)$ curve.

Based on the differences described above, it becomes obvious that the choice of the fitting strategy for unimodal fitting determines the fitting outcome for all soil hydraulic parameters, whereby especially the fitted α values showed significant differences.

After evaluating the hydraulic parameters of the unimodal fitting for different fitting strategies, we select only those soil samples ($N = 126$) that have been identified as bimodal by the lowest AIC amongst all fitting strategies and compare the estimated soil hydraulic parameters of the different fitting strategies. Here, it has to be noted that for the bimodal fit, in addition to the parameters θ_s and τ, only the α and n values from the matrix domain were used in the comparison, as the α and n values of the matrix domain differ greatly from those estimated for a unimodal model. Additionally, α and n values for the structural or macropore domain are highly uncertain, which is mainly caused by the low number of

TABLE 4 Statistics of α, n, θ_s, and τ of the unimodal and bimodal parameter for soil samples identified as bimodal by the lowest Akaike Information Criterion ($N = 126$)

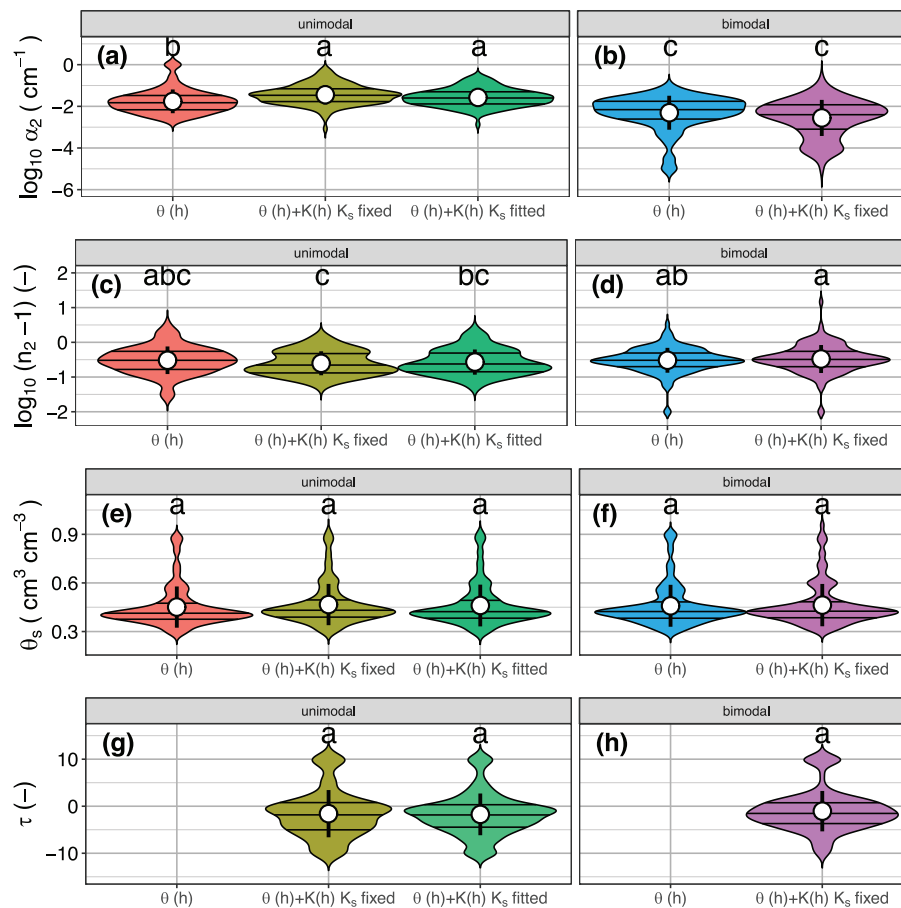|		Unimodal			Bimodal	
Parameter	Statistic	$\theta(h)$	$\theta(h) + K(h)$ K_s fixed	$\theta(h) + K(h)$ K_s fitted	$\theta(h)$	$\theta(h) + K(h)$ K_s fixed
α, cm^{-1}	Arithmetic mean	0.0174	0.0359	0.0265	0.0049	0.0028
	Median	0.0143	0.0352	0.0251	0.0074	0.0043
	Min.	0.0019	0.0008	0.0013	0.00001	0.00001
	Max.	1.0000	0.5680	0.2623	0.0802	0.2025
n	Arithmetic mean	1.460	1.345	1.395	1.414	1.558
	Median	1.298	1.207	1.221	1.298	1.329
	Min.	1.026	1.050	1.050	1.010	1.010
	Max.	4.473	2.607	3.486	4.410	15.627
θ_s, cm^3 cm^{-3}	Arithmetic mean	0.451	0.467	0.461	0.459	0.463
	Median	0.411	0.430	0.420	0.426	0.423
	Min.	0.289	0.299	0.296	0.297	0.300
	Max.	0.887	0.916	0.907	0.910	0.958
τ	Arithmetic mean	–	−1.583	−1.732	–	−1.048
	Median	–	−1.616	−1.549	–	−1.613
	Min.	–	−10.000	−10.000	–	−10.000
	Max.	–	10.000	10.000	–	10.000

Note. Values for α and n from the bimodal model are shown for the matrix domain. Values for α and n were log$_{10}$ transformed before calculating the mean and median values and then were anti-log transformed for calculating the presented values.

available data points in the wet region. This uncertainty has been partly addressed by Dettmann et al. (2014), who suggested fixing the n value for the macropore domain to a high value to overcome the uncertainty and to reduce the number of parameters to be fitted. The summaries of the statistics of the fitted α, n, θ_s, and τ are provided in Table 4.

Figure 3 shows the distribution of van Genuchten parameters for those soils identified as bimodal among the different fitting strategies. As can be seen, the α value is in generally larger for the unimodal fittings compared with the bimodal ones in terms of mean, median, the 25 and 75% quantile values (see Figures 3a,b), and the difference is also statistically significant, as all unimodal fittings differ from those based on the bimodal model. The significant difference is shown by alphabetical letters above each violin graph, whereby the different letters indicate differences between groups (fitting strategies). There is a significant difference between the fitting strategies in the case of estimating n as well (see Figures 3c,d). The largest mean and median values for n were detected for the bimodal $\theta(h) + K(h)$ K_s fixed strategy ($n = 1.558$ and 1.329) (Table 4). There is no significant difference between bimodal $\theta(h)$ and bimodal $\theta(h) + K(h)$ K_s fixed cases. The unimodal $\theta(h) + K(h)$ K_s fixed strategy has the lowest n values in terms of arithmetic mean and median values ($n = 1.345$ and 1.207, respectively). We also observe that the minimum and maximum values are affected by the choice of the fitting strategy, whereby the minimum values for both bimodal fitting strategies hit the lower bounds set to 1.01 and the max-

imum values for the bimodal $\theta(h) + K(h)$ K_s fixed strategy reached 15.627. For the estimated θ_s (see Figures 3e,f), no significant differences can be found between the different fitting strategies.

Differences in the estimated tortuosity (τ) are analyzed, whereby τ was only estimated for the case where both data $\theta(h)$ and $K(h)$ were used in the fitting. Table 4 shows that the estimated τ has significant variations with the minimum and maximum hitting the lower and upper bounds of the parameters (−10 and 10). The bimodal $\theta(h) + K(h)$ K_s fixed case showed higher τ in terms of arithmetic mean values ($\tau = -1.048$) and lower τ in terms of median values with $\tau = -1.613$ compared with the unimodal $\theta(h) + K(h)$ K_s fixed and $\theta(h) + K(h)$ K_s fitted cases. This is because the bimodal $\theta(h)+K(h)$ K_s fixed case produced more τ values hitting the upper bounds of 10 and less hitting the lower bounds of −10, which can be vividly shown in the violin plot for τ (Figure 3h). Again, both the mean and median values of τ deviated from the default of 0.5 but are close to the optimal value of −1 suggested by Schaap and Leij (2000). Rezanezhad et al. (2009) identified that tortuosity, porosity, and the hydraulic radius of the pores are the main aspects that influence $K(h)$. In this study, the soil porosity, being equivalent to the value of θ_s (soil porosity includes the dead-end pores compared with θ_s), has very close values in all fitting strategies. Compared with the fitting to $\theta(h)$-only strategy, fitting both $\theta(h)$ and $K(h)$ functions simultaneously for both the unimodal and bimodal cases greatly improved the

FIGURE 3 Violin plots for Mualem–van Genuchten α, n, and θ_s, and τ for soil samples identified as bimodal by the lowest Akaike Information Criterion ($N = 133$). The first three left cases (a, c, e, g) show parameters of unimodal fitting strategies, and the right two cases (b, d, f, h) are for bimodal. Centerline indicates the median, lower and upper line the 25th and 75th quantiles, the white point the arithmetic mean, and the vertical bars suggest the standard deviation. Note that for the bimodal samples, only the matrix α and n values of the bimodal model were used. Letters indicate significant differences ($p = .05$), whereby cases with different letters are significantly different

fitting by adding only tortuosity factor (τ), as indicated by the lowest AIC values, suggesting the importance of including tortuosity during the fitting. However, there are no significant differences between the fitting strategies based on τ neither on all samples nor on samples with bimodal characteristics.

The overall results indicate that the estimated α and n (or matrix α_2 and n_2 for bimodal fits) values are affected by the choice of the fitting strategy. In general, α is more influenced by structural porosity as already shown by Assouline and Or (2013), whereas n is known to be strongly determined by soil texture (e.g., Vereecken et al., 2010; Wösten et al., 2001). There are no significant differences between the estimated θ_s and τ for different strategies. By including τ, fitting both $\theta(h)$ and $K(h)$ functions simultaneously greatly improved the fitting for both the unimodal and bimodal cases.

Finally, both the mean and standard deviation of sand, clay fraction, bulk density, bimodal soil hydraulic parameters for both the macropore and matrix domains, and associated basic soil properties are shown in lookup tables in Supplemental

Tables S3 and S4. Soil samples identified to have bimodality by the lowest AIC and soil texture information are used for the statistical analysis. The soil textural information is available for all the Vereecken and UNSODA samples, whereas only very few EU-HYDI samples have textural information, leading to only 53 samples for the analysis.

3.4 | The α versus n relationship

3.4.1 | The oretical α vs. n relationship

To explore the relationship between α and n in the van Genuchten equation, we firstly investigate the relationship analytically (Equations 7–10). To this end, we calculated the derivative between α and n in order to match the mathematical relationship between S_e and h, which is shown graphically for a calculated water content at a given matric potential h (here $h = -200$ cm) in Figure 4. As can be seen, a clear
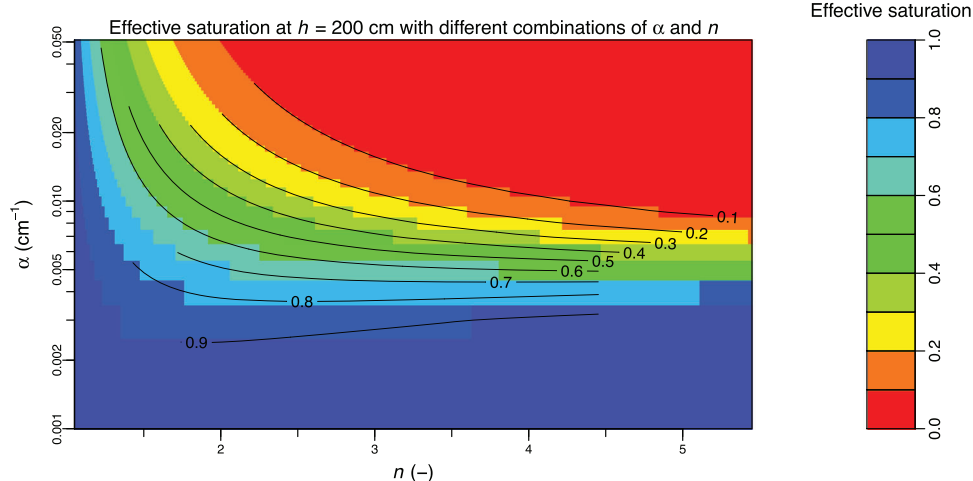
FIGURE 4 Calculated effective saturation (S_e) for different unimodal van Genuchten α (cm^{-1}) and n value for a given matric potential h of -200 cm

negative relationship between α and n for the dry regions (warm color in the figure) is detectable. At high saturation, α and n does not show a one-way relationship anymore, but illustrates a shift to a positive relationship when n is higher than $\sim$1.5, and a negative relationship when n is lower than $\sim$1.5. This is because the $(\alpha h)^n$ values are approximately the same order as one in the van Genuchten equation (see Equation 7) at high water contents.

3.4.2 | The α versus n relationship from literature

Besides the theoretical relationship between α and n discussed above, α and n parameters reported in the literature were collected and analyzed and shown in Figure 5. As can be seen, neither a linear positive nor a negative relationship can be detected from the entire data source. Additionally, some scattered data for α values ranging from 0.02 to 0.05 (cm^{-1}) with high values of n ($>$2.5) are reported (in red box). Those scattered data appear in almost all data sources and mostly belong to the USDA sand textural classification. Neglecting the high n values between α from 0.02 to 0.05, a negative trend can be assumed with a correlation coefficient $r = -0.42$ for all data sources. For a single data source, a weak negative trend between α and n can be identified by neglecting the outlier for the Rawls (red square cross), Puhlmann & Wilpert dataset (black diamond), and Botros (green star) dataset. However, neither any positive nor negative relationship can be detected for the Ahuja (red square), UNSODA dataset (orange triangle), and HYPRES (blue triangle) dataset.

Here, it has to be noted that the data from Rawls, Ahuja, UNSODA, and 'all' from Schaap and Leij (1998), as well as those from HYPRES represent mean values of the USDA soil textural classes and those averaging values among soil textu-
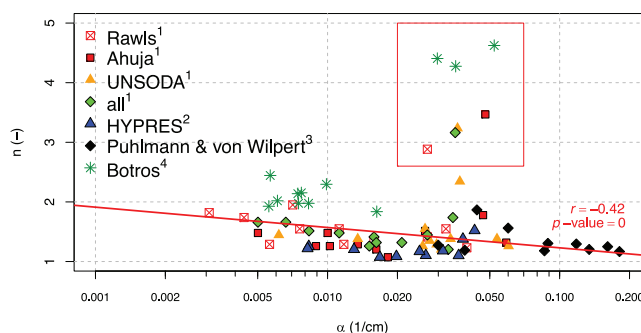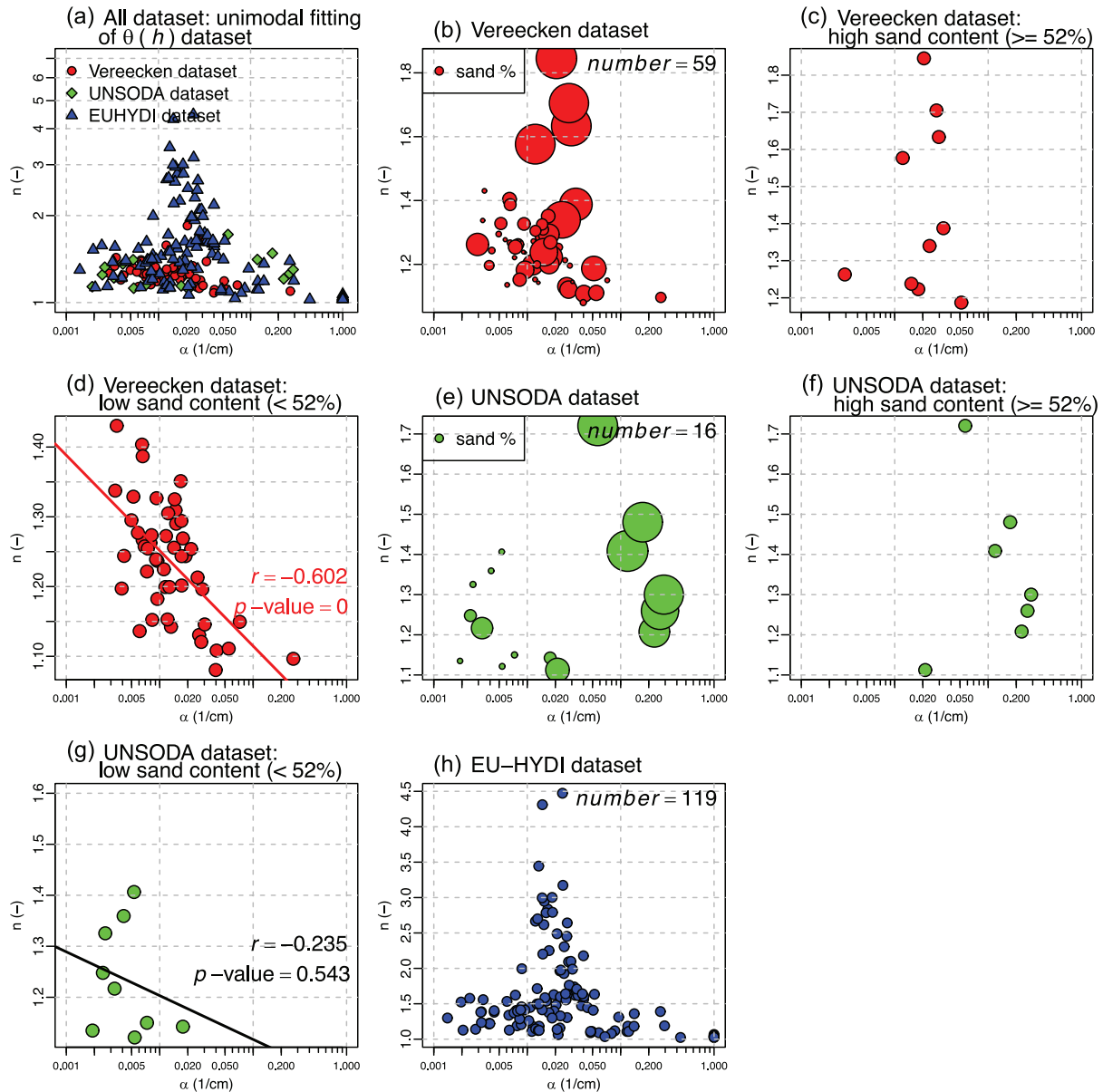

FIGURE 5 The relationship between van Genuchten α (cm^{-1}) vs. n taken from literature. The superscript in the legend indicate dataset from [1]Schaap and Leij (1998), [2]Wösten et al. (1999), [3]Puhlmann and von Wilpert (2012), and [4]Botros et al. (2009). Dataset from Schaap and Leij (1998) and Wösten et al. (1999) represent mean values for the USDA textural classes. The dataset with the α value in the range of .02 and .05 and high n values are encircled by the red box. The trendline was added for the remaining dataset excluding the data in the red box

ral classes might alleviate the relationship between α and n. In addition, the dataset shown in Figure 5 originates from different sources and is subject to different measurement methods, which might again relieve the relationship.

3.4.3 | The α versus n relationship in this study

After analyzing the α vs. n relationship from literature, the relationship from the dataset used in this study was also investigated, whereby all different fitting strategies for both the unimodal and bimodal models were used.

Figure 6a shows the α vs. n relationship from the unimodal fitting to $\theta(h)$ data only and all the three datasets used, whereby no obvious relationship can be deduced. Again, some

FIGURE 6 The relationship between van Genuchten α (cm^{-1}) vs. n from unimodal soil water retention curve (Equation 1) fitted to retention data only. (a) The case for all datasets including the Vereecken, unsaturated soil hydraulic database (UNSODA), and European hydropedological data inventory (EU-HYDI) dataset. (b), (c), (d) The cases for the entire, high-sand ($\geq$52%), and low-sand (<52%) content dataset from the Vereecken dataset. (e), (f), (g) The cases for the entire high-sand ($\geq$52%) and low-sand (<52%) content dataset from the UNSODA dataset; (h) The case from the EU-HYDI dataset

scatter (high values) for n with α values ranging from 0.01 to 0.1 can be detected as we have also seen in the literature data for α in the range from 0.02 to 0.05 (Figure 5). After dividing the data into high ($\geq$52%) and low (<52%) sand content, a negative correlation can be found between α and n in the case of low sand content (<52%) with a correlation coefficient of $r = -0.602$ and -0.235, respectively, for both Vereecken (Figure 6d) and UNSODA (Figure 6 g) datasets. The threshold value of 52% is the upper sand content limit of loam soils in the USDA soil texture triangle. Soil sand content >52% is classified as sandy loam, sandy clay loam, or similar tex-

ture, while the value <52% is categorized as loam or related texture. Other values ~52% were also investigated, which led to similar results. On the other hand, for samples with high sand content ($\geq$52%), a clear relationship cannot be identified (Figures 6c,f). Unfortunately, for the EU-HYDI samples, there was no possibility for the distinction of samples based on sand content because the dataset does not have affiliated soil particle size property for the samples selected for this analysis.

In the case of joint fitting of unimodal soil $\theta(h)$ and $K(h)$ data, the above-mentioned negative relationship is again detectable for the samples with low sand content with a

correlation coefficient of $r = -0.675$ and -0.775, respectively for Vereecken and UNSODA datasets (Supplemental Figures S1d,g). The correlation has been improved compared with the unimodal fitting to θ(h) data only because the joint fitting of soil θ(h) and K(h) data contained more information to constrain the α and n relationship. For samples with high sand content ($\geq$52%), similarly to the above case, a clear relationship is not visible.

For the α vs. n relationship obtained from the bimodal fitting to both θ(h) only and θ(h) + K(h) K_s fixed cases, no significant relationship can be observed between α vs. n of the macropore domain (See Supplemental Figures S2 and S3). This may result from the fact that the macropore domain is inherently challenging to characterize using classically used soil θ(h) and K(h) models, leading to bias or even incorrect parameters. The second reason is that even if the classically used models are appropriate to characterize the macropore domain, the macropore soil hydraulic parameters have no correlations intrinsically. The third reason is likely that the measurement of soil θ(h) and K(h) dataset does not have enough records in the sample near saturation to characterize the macropore domain, leading to high uncertainty of the fitted parameters.

For the matrix domain from the bimodal fitting to θ(h) data, the two parameters are also negatively correlated for the low sand content samples (<52%) with correlation coefficient $r = -0.658$ and -0.763, respectively, for the Vereecken and UNSODA databases by excluding the extremely low α values (0.00005 cm^{-1}) (Figures 7b,e). Those extremely low α values hit the lower threshold set for α in the parameter estimation, suggesting the optimal α values may not be found. For the EU-HYDI samples, the negative correlation between the two parameters is not as clear, and some scatter with high n values in the range of 0.001 to 0.1 for α exist. We observed a similar relationship for the matrix domain from the bimodal fitting to θ(h) + K(h) K_s fixed case in Supplemental Figure S4.

The α parameter is related to the scaling of the inverse of air entry pressure, indicating the size of the largest pores in the soil porous system. For the low sand content situation (sand fraction <52%), fine-textured samples may have connected large pores, leading to low air entry pressure and therefore to high α values. For the silt-sized samples, as we mentioned in the General Database Statistics and Properties section that silt-sized fraction sometimes shows characteristics similar to both sand and clay-sized constituents, it may show small-sized pores and therefore low α values. n, on the other hand, indicates the width of the pore size distribution, suggesting the slope of the soil θ(h) and K(h) curves. Fine-textured soils typically have low n values, suggesting a broad range of pore size distributions and a gradual increasing slope in θ(h) curves. As n becomes larger, the pore size distribution becomes narrower and the slope of soil θ(h) curves becomes steeper. This leads to the negative relationship between α and n for low sand con-

tent samples (sand fraction <52%). The relationship between α and n in our analysis contradicts the classical notion that there is a positive relationship between α and n, which was often based on mean values for textural classes to analyze the dataset and the mean values may weaken the relationship. In addition, the entire soil textural classes were often analyzed, which may produce the artifact that α and n are positively correlated.

3.5 | The relationship between bimodal parameters and other soil properties

The above analysis showed the relationship between α and n both in analytical solutions and in a statistical approach from literature data and the dataset used in this study. We also analyzed whether the ratios of macropore and matrix soil hydraulic parameters describing the shape of the θ(h) curve correlate with the weighting factor (w, being used to superpose the macropore and matrix domain pore systems, see Equations 3 and 4) and basic soil properties. The presence of correlation amongst these parameters and soil properties may pave a way forward in the PTF development to estimate bimodal soil hydraulic properties so that end-users can derive bimodal soil hydraulic properties from readily available basic soil properties.

For the bimodal fitting to θ(h) only, Figure 8a shows that the ratio of α between macropore and matrix domain [$\log(\alpha_1/\alpha_2)$] has a moderate positive correlation with the weighting factor (w) for the Vereecken and UNSODA dataset, with a correlation coefficient (r) of 0.504 and 0.691, respectively, while the correlation for the EU-HYDI dataset shows a negative relationship with r of -0.410. It has to be noted that we removed the parameters that hit the upper and lower parameter bounds and also parameters that were fitted to the same values for the macropore and matrix domains, as those parameters might be unreliable. Unfortunately, the former two datasets and the latter one shows contradicting trends. To the best of our knowledge, the trend from the Vereecken and UNSODA datasets seems more reasonable. The reason is that, firstly, in contrast to the bimodal fitting to θ(h) only, the bimodal fitting to θ(h) + K(h) K_s fixed case shown in Supplemental Figure S5a also indicates that all three databases have a positive relationship between [$\log(\alpha_1/\alpha_2)$] and the weighting factor (w). In contrast, the EU-HYDI dataset has an even higher positive correlation (r = 0.648) than the results from the Vereecken and UNSODA datasets. The bimodal fitting to θ(h) + K(h) K_s fixed case used both θ(h) and K(h) dataset to fit the bimodal soil hydraulic properties and may provide additional information to constrain the fitted parameters and therefore may provide improved fitted parameters. Secondly, the higher weighting factor (w) suggests that the macropore domain
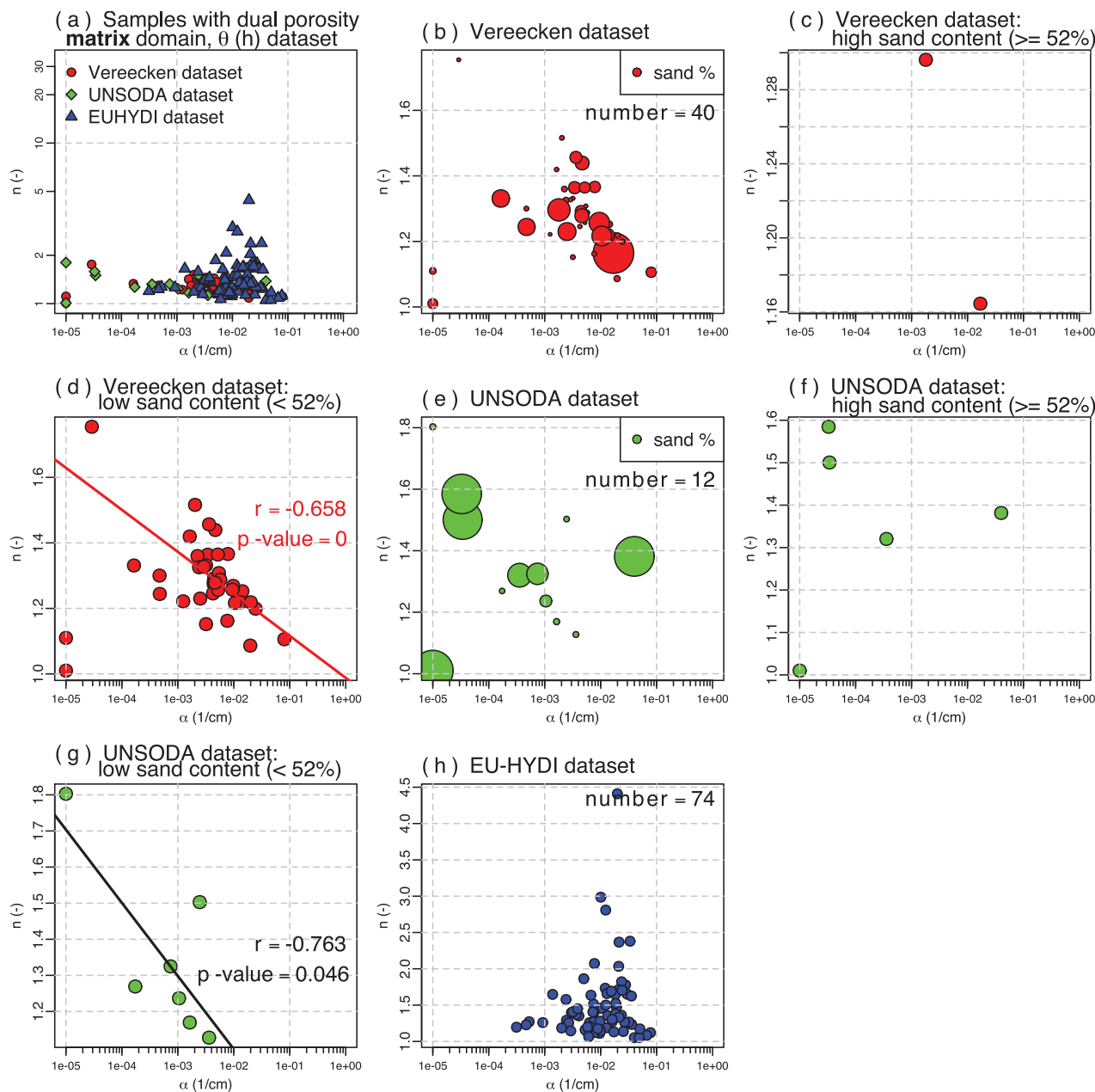
FIGURE 7 The relationship between van Genuchten α (cm^{-1}) versus n (−) of matrix domain from the bimodal fitting of soil water retention curve (Equation 3) to the retention dataset. (a) The case for samples with dual porosity for all three datasets including the Vereecken, unsaturated soil hydraulic database (UNSODA), and European hydropedological data inventory (EU-HYDI). (b), (c), (d) The cases for the entire high-sand (≥52%) and low-sand (< 52%) content dataset from the Vereecken dataset. (e), (f) (g) The cases for the entire, high-sand (≥52%) and low-sand (<52%) content dataset from the UNSODA dataset. (h) The case from the EU-HYDI dataset

is more dominated, whereas related higher α_1/α_2 ratios are caused by larger α_1 and smaller α_2 values, being associated with larger macropores and finer-textured matrix, indicating larger macropores embedded in a relatively fine matrix. This is consistent with the positive correlation observed from the Vereecken and UNSODA dataset. Therefore, we conclude that by including the $K(h)$ data in the fitting, the EU-HYDI dataset improves the fitting results shown in Supplemental Figure S5a.

For the bimodal fitting to $\theta(h)$ only, we also observed a strong positive correlation between the ratio of α among macropore and matrix domain [log(α_1/α_2)] with soil sand content (sand percentage) for the UNSODA dataset with an $r = 0.965$ in Figure 8c, while the correlation for the Vereecken dataset is very weak. The positive correlation suggests that soil samples with a high sand content are likely to have a high ratio of α values between macropore and matrix domain [log(α_1/α_2)], indicating large macropores (α_1) embedded in a
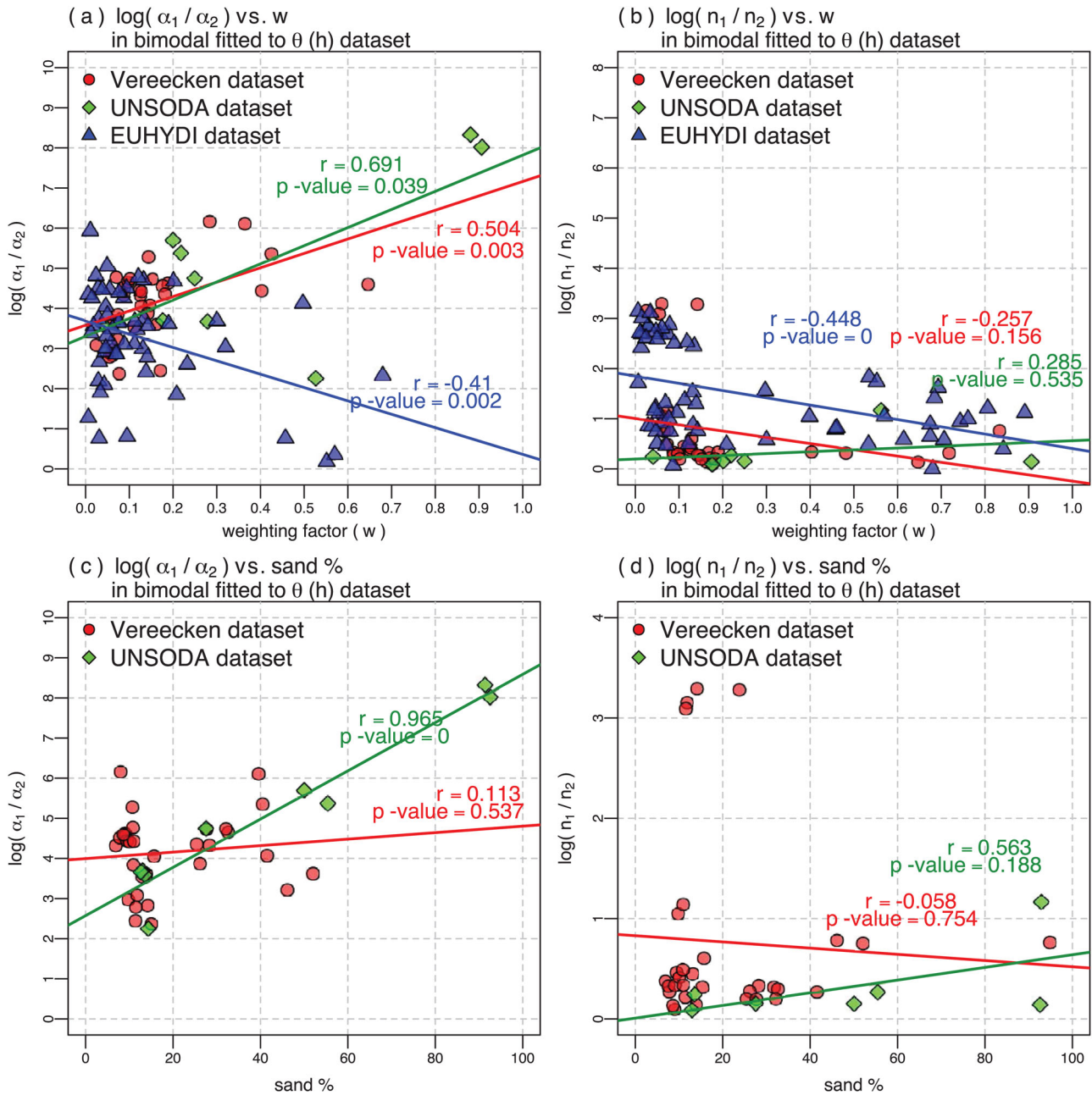
FIGURE 8 van Genuchten (a) α and (b) n in the corresponding macropore (α_1 or n_1) and matrix (α_2 or n_2) domain vs. weighting factors (w). (c) and (d) corresponding ratios of macropore (α_1 or n_1) and matrix (α_2 or n_2) domain vs. soil sand content. The parameters were obtained based on bimodal fitting to the retention dataset (Equation 3). Values that hit the upper and lower parameter bounds and that macropore and matrix parameters that produced the same values were removed

fine matrix (α_2). The same observations and conclusion were also drawn for the bimodal fitting to the $\theta(h) + K(h)$ K_s fixed case shown in Supplemental Figure S5c.

Because the bimodal fitting to $\theta(h)$ from the EU-HYDI dataset is questionable in Figure 8a, we focus on the Vereecken and UNSODA datasets to analyze the relationship between the ratio of n in the macropore and matrix domain [$\log(n_1/n_2)$] with the weighting factors (w), which shows that the correlation is not apparent in Figure 8b. We derived the same conclusion for the relationship between $\log(n_1/n_2)$ and

w based on the bimodal fitting to $\theta(h) + K(h)$ K_s fixed case shown in Supplemental Figure S5b. The unclear correlation may result from the fitted macropore n_1 value having very large uncertainty. Dettmann et al. (2014) even suggested that the macropore n_1 parameter can be set to the upper parameter limit for the largest macropores during the fitting, and a further increase of the upper limits would not significantly improve model performance but often lead to a raised instability of the numerical solution when those data are used in predictive models. The correlation between $\log(n_1/n_2)$ and

sand content (sand percentage) for the UNSODA dataset has an $r = 0.563$ in Figure 8d. However, $r = -0.058$ for the Vereecken dataset, suggesting no correlation between the two quantities. The bimodal fitting to $\theta(h) + K(h)$ K_s fixed case is shown in Supplemental Figure S5d and also shows a very weak correlation of $r = 0.246$ for the Vereecken dataset. We attribute this again to the large uncertainty of the fitted macropore n_1 parameter.

The ratio of soil hydraulic properties between the macropore and matrix domain with the other basic soil properties—soil bulk density and soil organic carbon content from the dataset—is also investigated, and we did not find an apparent tendency. The correlation graphs are, therefore, not shown.

4 | CONCLUSIONS

In this study, we collected and established a database of basic soil properties and soil hydraulic properties for analyzing the effect of soil structure and macropore on soil $\theta(h)$ and $K(h)$ characteristics. To this end, both unimodal and bimodal soil hydraulic functions were used to describe the soil hydraulic properties. Our work leads to the following major conclusions:

1. We analyzed three databases (Vereecken, UNSODA, and EU-HYDI) that have been used to develop unimodal PTFs to study the presence of bimodality in measured soil $\theta(h)$ and $K(h)$ curves. Out of a total of 19,510 samples, only 194 samples were suitable to perform this analysis, as these samples were undisturbed and contained enough data points in the wet range of both $\theta(h)$ and $K(h)$. More than half of them (65%) showed the presence of bimodality independent of the textural class. We found that bimodality is therefore not only limited to fine-textured soils, for which bimodal models have been shown to provide good results in describing the $\theta(h)$ and $K(h)$ curves, but it is also observed in many coarser-textured soils. Silt loam soil samples, accounting for the largest portion of soil samples containing soil texture information in our database, were best fitted by the bimodal fit. The fact that only 1% of the data in three databases together was suitable for analyzing the presence of bimodality and thus structural effects on soil hydraulic properties calls for a community effort to jointly analyzing $\theta(h)$ and $K(h)$ on undisturbed samples. The samples should be measured in a standardized manner and provide auxiliary variables beyond soil texture, bulk density, and organic matter, which allows for the development of PTFs that account for the effect of bimodality or even multimodality on soil hydraulic properties.
2. The classical approach that uses either fitted or PTF-estimated retention parameters of the van Genuchten model in combination with a fixed tortuosity parameter of

0.5 and a measured K_s value leads to poor performance in estimating the unsaturated $K(h)$. Only one sample out of a total of 194 samples was best fitted by using this approach and showed a good prediction of unsaturated $K(h)$. Yet, this is still the current practice in estimating unsaturated $K(h)$, and it shows that the standard MvG model is not suitable for predicting unsaturated $K(h)$ solely from $\theta(h)$ data and measured K_s. Fitting both unimodal $\theta(h)$ and unsaturated $K(h)$ functions simultaneously significantly outperforms the fitting of $\theta(h)$ data only and then using the MvG parameters and K_s to calculate the $K(h)$. Fitting based on bimodal $\theta(h) + K(h)$ functions simultaneously also outperformed the fitting based on bimodal only $\theta(h)$ and using MvG parameters and K_s to calculate $K(h)$. Our study further suggests that soil hydraulic characteristics that show the presence of structural properties cannot be appropriately described with unimodal models, and more accurate fitting can only be performed by considering dual-porosity models. Therefore, including the presence of soil structural effects on soil hydraulic parameters may improve the development of PTFs, especially for PTFs estimating the $K(h)$ function.
3. For the unimodal model, α values (scaling parameter related to the inverse of air entry pressure) in the van Genuchten model showed systematic differences between different fitting strategies. For the bimodal model, the α and n (a measure for the pore size distribution) estimated for the matrix domain show significant differences among different fitting strategies. In contrast, saturated water content (θ_s) and tortuosity factor (τ) did not show significant differences, suggesting that both α and n values are more influenced by structural porosity.
4. We analyzed the relationship between α vs. n analytically from literature data and from our established dataset. For the analytical method, we showed in the van Genuchten model that when $m = 1 - 1/n$, there is a clear negative relationship between α and n for the dry regions. For the near-saturation regions, the relationship between α and n depends upon the n parameter. The literature dataset, partly providing textural class averaged parameters, does not show a clear relationship between α and n. We suspect that the averaging of the values might alleviate the relationship between the two parameters.
5. Our established dataset showed a clear negative relationship between α and n parameters in the van Genuchten model in the case of low sand content (<52%) both for all the fitting cases [i.e., unimodal fitting to $\theta(h)$ data, joint fitting of unimodal $\theta(h) + K(h)$ data, the matrix domain properties of the bimodal fitting to $\theta(h)$ data, and the matrix domain properties of the bimodal fitting to $\theta(h) + K(h)$ data] from the Vereecken and UNSODA dataset. This negative trend seems to contradict the classical notion that larger α values usually correspond to larger n values.

Because the samples selected from the EU-HYDI dataset do not have associated soil particle size information for most data selected, it is impossible to analyze the apparent relationship between α and *n* for the case of low sand content. In addition, no significant relationship was observed for the two parameters of the macropore domain.

6. The ratio of α between macropore and matrix domain [$\log(\alpha_1/\alpha_2)$] showed a positive correlation with the weighting factor for most of the analyzed samples, which is used to describe how much the macropore domain is accounted for in the entire domain. The ratio of α between the two domains is also observed to have a positive relationship with the soil sand content (sand percentage) especially for the UNSODA dataset with a correlation coefficient of 0.965. This relationship is promising and paves a way forward the PTF development of bimodal soil hydraulic properties.

AUTHOR CONTRIBUTIONS

Yonggen Zhang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing – original draft; Writing – review & editing. Lutz Weihermüller: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing – original draft; Writing – review & editing. Brigitta Toth: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing – original draft; Writing – review & editing. Muhammad Noman: Formal analysis; Investigation. Harry Vereecken: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ORCID

Yonggen Zhang https://orcid.org/0000-0001-9242-2558
Harry Vereecken https://orcid.org/0000-0002-8051-8517

REFERENCES

Addiscott, T. M. (1977). A simple computer model for leaching in structured soils. *European Journal of Soil Science*, *28*, 554–563. https://doi.org/10.1111/j.1365-2389.1977.tb02263.x

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Ardia, D., Mullen, K. M., Peterson, B. G., & Ulrich, J. (2020). *'DEoptim': Differential evolution in 'R'* (v2.2-5). https://cran.r-project.org/web/packages/DEoptim/DEoptim.pdf

Assouline, S., & Or, D. (2013). Conceptual and parametric representation of soil hydraulic properties: A review. *Vadose Zone Journal*, *12*, 1–20. https://doi.org/10.2136/vzj2013.07.0121

Bendtsen, C. (2012). *pso: Particle swarm optimization* (R package v1.0.3). https://cran.r-project.org/package=pso

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beven, K., & Germann, P. (1981). Water flow in soil macropores II. A combined flow model. *Journal of Soil Science*, *32*, 15–29. https://doi.org/10.1111/j.1365-2389.1981.tb01682.x

Botros, F. E., Harter, T., Onsoy, Y. S., Tuli, A., & Hopmans, J. W. (2009). Spatial variability of hydraulic properties and sediment characteristics in a deep alluvial unsaturated zone. *Vadose Zone Journal*, *8*, 276–289. https://doi.org/10.2136/vzj2008.0087

Brewer, R. (1964). *Fabric and mineral analysis of soils*. J. Wiley & Sons.

Brooks, R., & Corey, A. (1964). *Hydraulic properties of porous media*. Colorado State University. https://mountainscholar.org/bitstream/handle/10217/61288/HydrologyPapers_n3.pdf

Clapp, R. B., & Hornberger, G. M. (1978). Empirical equations for some soil hydraulic properties. *Water Resources Research*, *14*, 601–604. https://doi.org/10.1029/WR014i004p00601

Coppola, A. (2000). Unimodal and bimodal descriptions of hydraulic properties for aggregated soils. *Soil Science Society of America Journal*, *64*, 1252–1262. https://doi.org/10.2136/sssaj2000.6441252x

Coppola, A., Basile, A., Comegna, A., & Lamaddalena, N. (2009). Monte Carlo analysis of field water flow comparing uni-and bimodal effective hydraulic parameters for structured soil. *Journal of Contaminant Hydrology*, *104*, 153–165. https://doi.org/10.1016/j.jconhyd.2008.09.007

Coppola, A., Comegna, V., Basile, A., Lamaddalena, N., & Severino, G. (2009). Darcian preferential water flow and solute transport through bimodal porous systems: Experiments and modelling. *Journal of Contaminant Hydrology*, *104*, 74–83. https://doi.org/10.1016/j.jconhyd.2008.10.004

De Mendiburu, F. (2014). *Agricolae: statistical procedures for agricultural research* (R package v1). https://cran.r-project.org/web/packages/agricolae/index.html

Dettmann, U., Bechtold, M., Frahm, E., & Tiemeyer, B. (2014). On the applicability of unimodal and bimodal van Genuchten–Mualem based models to peat and other organic soils under evaporation conditions. *Journal of Hydrology*, *515*, 103–115. https://doi.org/10.1016/j.jhydrol.2014.04.047

Dexter, A. R. (1988). Advances in characterization of soil structure. *Soil and Tillage Research*, *11*, 199–238. https://doi.org/10.1016/0167-1987(88)90002-5

Dexter, A. R., Czyż, E. A., Richard, G., & Reszkowska, A. (2008). A user-friendly water retention function that takes account of the textural and structural pore spaces in soil. *Geoderma*, *143*, 243–253. https://doi.org/10.1016/j.geoderma.2007.11.010

Dexter, A. R., & Richard, G. (2009). The saturated hydraulic conductivity of soils with *n*-modal pore size distributions. *Geoderma*, *154*, 76–85. https://doi.org/10.1016/j.geoderma.2009.09.015

Durner, W. (1994). Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resources Research*, *30*, 211–223. https://doi.org/10.1029/93WR02676

Gerke, H. H., & van Genuchten, M. T. (1993). A dual porosity model for simulating the preferential movement of water and solutes in structured porous media. *Water Resources Research*, *29*, 305–319. https://doi.org/10.1029/92WR02339

Germann, P., & Beven, K. (1981). Water flow in soils macropores I: An experimental approach. *Journal of Soil Science*, *32*, 1–13. https://doi.org/10.1111/j.1365-2389.1981.tb01681.x

Hadas, A. (1987). Long-term tillage practice effects on soil aggregation modes and strength. *Soil Science Society of America Journal*, *51*, 191–197. https://doi.org/10.2136/sssaj1987.03615995005100010040x

Haghverdi, A., Öztürk, H. S., & Durner, W. (2020). Studying unimodal, bimodal, PDI and bimodal-PDI variants of multiple soil water retention models: II. Evaluation of parametric pedotransfer functions against direct fits. *Water*, *12*, 896. https://doi.org/10.3390/w12030896

Jadoon, K. Z., Weihermüller, L., Scharnagl, B., Kowalsky, M. B., Bechtold, M., Hubbard, S. S., Vereecken, H., & Lambot, S. (2012). Estimation of soil hydraulic parameters in the field by integrated hydrogeophysical inversion of time-lapse ground-penetrating radar data. *Vadose Zone Journal*, *11*, vzj2011.0177. https://doi.org/10.2136/vzj2011.0177

Jarvis, N. J. (2007). A review of non-equilibrium water flow and solute transport in soil macropores: Principles, controlling factors and consequences for water quality. *European Journal of Soil Science*, *58*, 523–546. https://doi.org/10.1111/j.1365-2389.2007.00915.x

Jarvis, N. J., Jansson, P.-E., Dik, P. E., & Messing, I. (1991). Modelling water and solute transport in macroporous soil, I. Model description and sensitivity analysis. *European Journal of Soil Science*, *42*, 59–70. https://doi.org/10.1111/j.1365-2389.1991.tb00091.x

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks* (pp. 1942–1948). IEEE.

Lal, R., & Shukla, M. K. (2004). *Principles of soil physics*. CRC Press.

Lehmann, P., Bickel, S., Wei, Z., & Or, D. (2020). Physical constraints for improved soil hydraulic parameter estimation by pedotransfer functions. *Water Resources Research*, *56*, e2019WR025963. https://doi.org/10.1029/2019WR025963

Li, X., Li, J. H., & Zhang, L. M. (2014). Predicting bimodal soil–water characteristic curves and permeability functions using physically based parameters. *Computers and Geotechnics*, *57*, 85–96. https://doi.org/10.1016/j.compgeo.2014.01.004

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *12*, 513–522. https://doi.org/10.1029/WR012i003p00513

Mullen, K. M., Ardia, D., Gil, D. L., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, *40*, 1–26. https://doi.org/10.18637/jss.v040.i06

Nemes, A., Schaap, M. G., Leij, F. J., & Wösten, J. H. M. (2001). Description of the unsaturated soil hydraulic database UNSODA version 2.0. *Journal of Hydrology*, *251*, 151–162. https://doi.org/10.1016/S0022-1694(01)00465-6

Oades, J. M., & Waters, A. G. (1991). Aggregate hierarchy in soils. *Australian Journal of Soil Research*, *29*, 815–828. https://doi.org/10.1071/SR9910815

Othmer, H., Diekkrüger, B., & Kutilek, M. (1991). Bimodal porosity and unsaturated hydraulic conductivity. *Soil Science*, *152*, 139–149. https://doi.org/10.1097/00010694-199109000-00001

Peters, A., Durner, W., & Wessolek, G. (2011). Consistent parameter constraints for soil hydraulic functions. *Advances in Water Resources*, *34*, 1352–1365. https://doi.org/10.1016/j.advwatres.2011.07.006

Philip, J. R. (1968). The theory of absorption in aggregated media. *Australian Journal of Soil Research*, *6*, 1–19. https://doi.org/10.1071/SR9680001

Price, K. V., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution—A practical approach to global optimization*. Springer-Verlag.

Priesack, E., & Durner, W. (2006). Closed-form expression for the multimodal unsaturated conductivity function. *Vadose Zone Journal*, *5*, 121–124. https://doi.org/10.2136/vzj2005.0066

Puhlmann, H., & von Wilpert, K. (2012). Pedotransfer functions for water retention and unsaturated hydraulic conductivity of forest soils. *Journal of Plant Nutrition and Soil Science*, *175*, 221–235. https://doi.org/10.1002/jpln.201100139.221

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rezanezhad, F., Quinton, W. L., Price, J. S., Elrick, D., Elliot, T. R., & Heck, R. J. (2009). Examining the effect of pore size distribution and shape on flow through unsaturated peat using computed tomography. *Hydrology and Earth System Sciences*, *13*, 1993–2002. https://doi.org/10.5194/hess-13-1993-2009

Ross, P. J., & Smettem, K. R. (1993). Describing soil hydraulic properties with sums of simple functions. *Soil Science Society of America Journal*, *57*, 26–29. https://doi.org/10.2136/sssaj1993.03615995005700010006x

Schaap, M. G., & Leij, F. J. (1998). Database related accuracy and uncertainty of pedotransfer functions. *Soil Science*, *163*, 765–779. https://doi.org/10.1097/00010694-199810000-00001

Schaap, M. G., & Leij, F. J. (2000). Improved prediction of unsaturated hydraulic conductivity with the Mualem–van Genuchten model. *Soil Science Society of America Journal*, *64*, 843–851. https://doi.org/10.2136/sssaj2000.643843x

Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (2001). Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, *251*, 163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Schaap, M. G., & van Genuchten, M. T. (2006). A modified Mualem–van Genuchten formulation for improved description of the hydraulic conductivity near saturation. *Vadose Zone Journal*, *5*, 27–34. https://doi.org/10.2136/vzj2005.0005

Schwen, A., Zimmermann, M., & Bodner, G. (2014). Vertical variations of soil hydraulic properties within two soil profiles and its relevance for soil water simulations. *Journal of Hydrology*, *516*, 169–181. https://doi.org/10.1016/j.jhydrol.2014.01.042

Schuh, W. M., & Cline, R. L. (1990). Effect of soil properties on unsaturated hydraulic conductivity pore-interaction factors. *Soil Science Society of America Journal*, *54*, 1509–1519. https://doi.org/10.2136/sssaj1990.03615995005400060001x

Scotter, D. R. (1978). Preferential solute movement through larger soil voids. I. Some computations using simple theory. *Australian Journal of Soil Research*, *16*, 257–267. https://doi.org/10.1071/SR9780257

Smettem, K. R. J., & Kirkby, C. (1990). Measuring the hydraulic properties of a stable aggregated soil. *Journal of Hydrology*, *117*, 1–13. https://doi.org/10.1016/0022-1694(90)90084-B

Spohrer, K., Herrmann, L., Ingwersen, J., & Stahr, K. (2006). Applicability of uni- and bimodal retention functions for water flow modeling in a tropical Acrisol. *Vadose Zone Journal*, *5*, 48–58. https://doi.org/10.2136/vzj2005.0047

Szabó, B., Weynants, M., & Weber, T. K. (2021). Updated European hydraulic pedotransfer functions with communicated uncertainties in the predicted variables (euptfv2). *Geoscientific Model Development*, *14*, 151–175. https://doi.org/10.5194/gmd-14-151-2021

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., & Tóth, G. (2015). New generation of hydraulic pedotransfer functions for Europe. *European Journal of Soil Science*, *66*, 226–238. https://doi.org/10.1111/ejss.12192

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, *44*, 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y. A., Padarian, J., Schaap, M. G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M. J., Weihermüller, L., Zacharias, S., Zhang, Y., & Vereecken, H. (2017). Pedotransfer functions in Earth system science: Challenges and perspectives. *Reviews of Geophysics*, *55*, 1199–1256. https://doi.org/10.1002/2017RG000581

Vereecken, H., Maes, J., & Feyen, J. (1990). Estimating unsaturated hydraulic conductivity from easily measured soil properties. *Soil Science*, *149*, 1–12. https://doi.org/10.1097/00010694-199001000-00001

Vereecken, H., Maes, J., Feyen, J., & Darius, P. (1989). Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content. *Soil Science*, *148*, 389–403. https://doi.org/10.1097/00010694-198912000-00001

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M. G., & van Genuchten, M. T. (2010). Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: A review. *Vadose Zone Journal*, *9*, 795–820. https://doi.org/10.2136/vzj2010.0045

Weber, T. K. D., Finkel, M., da Conceição Gonçalves, M., Vereecken, H., & Diamantopoulos, E. (2020). Pedotransfer function for the Brunswick soil hydraulic property model and comparison to the van Genuchten–Mualem model. *Water Resources Research*, *56*, e2019WR026820. https://doi.org/10.1029/2019WR026820

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. *Journal of Advances in Modeling Earth Systems*, *13*, e2020MS002404. https://doi.org/10.1029/2020MS002404

Weynants, M., Montanarella, L., Tóth, G., Arnoldussen, A., Anaya Romero, M., Bilas, G., Børresen, T., Cornelis, W., Daroussin, J., Gonçalves, M. D. C., Haugen, L.-E., Hennings, V., Houšková, B., Iovino, M., Javaux, M., Keay, C. A., Kätterer, T., Kværnø, S., Laktinova, T., … Wösten, H. (2013). *European hydropedological data inventory (EU-HYDI)*. Joint Research Centre, Institute for Prospective Technological Studies, Publications Office. https://data.europa.eu/doi/10.2788/5936

Weynants, M., Vereecken, H., & Javaux, M. (2009). Revisiting Vereecken pedotransfer functions: Introducing a closed-form hydraulic model. *Vadose Zone Journal*, *8*, 86–95. https://doi.org/10.2136/vzj2008.0062

Wösten, J. H. M., Lilly, A., Nemes, A., & Le Bas, C. (1999). Development and use of a database of hydraulic properties of European soils. *Geoderma*, *90*, 169–185. https://doi.org/10.1016/S0016-7061(98)00132-3

Wösten, J. H. M., Pachepsky, Y. A., & Rawls, W. J. (2001). Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*, *251*, 123–150. https://doi.org/10.1016/S0022-1694(01)00464-4

Xiang, Y., Gubian, S., Suomela, B., & Hoeng, J. (2013). Generalized simulated annealing for global optimization: The GenSA package. *R Journal*, *5*, 13–28.

Zhang, Y., & Schaap, M. G. (2017). Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology*, *547*, 39–53. https://doi.org/10.1016/j.jhydrol.2017.01.004

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.