AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR039337

Soil Physics-Informed Neural Networks to Estimate Bimodal Soil Hydraulic Properties

Jieliang Zhou[1,2], Yunquan Wang[2], Pengfei Qi[2], Rui Ma[2], Harry Vereecken[3], Budiman Minasny[4], and Yonggen Zhang[5]

[1]Institute of Geological Survey, China University of Geosciences, Wuhan, China, [2]Key Laboratory of Groundwater Quality and Health (China University of Geosciences), School of Environmental Studies, Ministry of Education, China University of Geosciences at Wuhan, Wuhan, PR China, [3]Agrosphere Institute, IBG-3, Forschungszentrum Jülich GmbH, Jülich, Germany, [4]School of Life and Environmental Sciences & Sydney Institute of Agriculture, The University of Sydney, Sydney, NSW, Australia, [5]School of Earth System Science, Tianjin University, Tianjin, China

Key Points:

- Novel Pedotransfer functions (PTFs) were developed to capture bimodal soil hydraulic properties
- The new PTFs effectively captured bimodality in hydraulic conductivity curves
- PTFs developed using soil physics-informed neural networks significantly outperformed traditional PTFs developed with machine learning

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Wang,
wangyq@cug.edu.cn

Citation:

Zhou, J., Wang, Y., Qi, P., Ma, R., Vereecken, H., Minasny, B., & Zhang, Y. (2025). Soil Physics-Informed Neural Networks to estimate bimodal soil hydraulic properties. *Water Resources Research*, *61*, e2024WR039337. https://doi.org/10.1029/2024WR039337

Received 31 OCT 2024
Accepted 19 SEP 2025

Abstract Pedotransfer functions (PTFs) are widely used to estimate soil hydraulic properties (SHPs) from easily measurable soil properties. However, most existing PTFs are based on unimodal hydraulic models, which fail to capture the bimodal behavior of hydraulic properties caused by soil structure. In this study, we developed new PTFs using two bimodal hydraulic models and introduced a soil physics–informed neural network to embed these models into the training process. The results showed that the new PTFs effectively represented bimodality in hydraulic conductivity curves, achieving a root mean square error of 0.531(K in cm/d) on the test set, compared with 0.626 for unimodal models. They also improved predictions of soil water retention curves but struggled with bimodal cases for some samples, likely due to the limited number of bimodal retention curves in the training data set. Evaluation on an independent data set showed that the error for hydraulic conductivity predicted by the new functions was about one-third that of conventional approaches. In addition, the proposed soil physics–informed neural network, which directly optimizes SHPs, outperformed the conventional approaches that optimize fitted model parameters. We also found that whether water retention and hydraulic conductivity are optimized separately or simultaneously has a large effect on performance. Nonetheless, the lack of explicit soil structure information in the input data, along with limited measurements near saturation, continues to constrain accuracy. This emphasizes the need to develop a more comprehensive soil hydraulic database.

1. Introduction

The accurate determination of soil hydraulic properties (SHPs), which include the soil water retention curve (SWRC) and the hydraulic conductivity curve (HCC), is essential for understanding soil water flow and biogeochemical processes across various scales (e.g., Netto et al., 1999; Wang et al., 2018; Youngs, 1988). However, direct measurements of SHPs are costly, time-consuming, and limited by scale and location, making them impractical for large-scale applications (Castellini et al., 2021). Consequently, pedotransfer functions (PTFs), which estimate SHPs from easily measured soil properties such as soil texture, have become a standard method for SHPs estimation (Van Looy et al., 2017; Weber et al., 2024).

In response to the increasing demand for SHP data for global land surface and soil processes modeling, PTFs development has advanced rapidly in recent decades, yielding more precise and reliable SHP estimates (e.g., Botula et al., 2014; McBratney et al., 2002; Patil & Singh, 2016; Van Looy et al., 2017; Vereecken et al., 2022; Weber et al., 2024). Efforts to enhance PTF performance include:

1. Expanding input variables beyond soil properties (e.g., Chirico et al., 2010; Pachepsky et al., 2006; Szabó et al., 2021). For example, S. Gupta et al. (2021) incorporated high-resolution remote sensing data, such as climate and vegetation indices, to improve predictions of saturated hydraulic conductivity (K_s).
2. Employing more sophisticated soil hydraulic models for PTF development (Rudiyanto et al., 2021; Wang, Zhou, et al., 2022). For instance, Wang, Zhou, et al. (2022) used a novel soil hydraulic model that accounts for both capillarity and adsorption forces, outperforming the traditional capillarity-based models, particularly under dry conditions.
3. Utilizing more efficient and accurate algorithms to establish relationships between soil properties and SHPs, evolving from linear regression (S. C. Gupta & Larson, 1979) to machine learning (Jalabert et al., 2010;

Pachepsky et al., 1996; Wang, Zhou, et al., 2022) and to Soil Physics-Informed Neural Networks (SPINN) (Minasny & McBratney, 2002; Norouzi et al., 2025; Rudiyanto et al., 2021).

Here, the term SPINN refers to the integration of physics knowledge, such as soil hydraulic models, into neural networks. Although existing soil hydraulic models are partly empirical—particularly in the formulation of the SWRC—they also incorporate physical principles in the development of the HCC. For example, the capillary theory is commonly applied to derive the hydraulic conductivity function (Mualem, 1976). SPINN-based PTFs directly link inputs to measured SHPs, and unlike conventional methods, which link inputs to fitted soil hydraulic parameters, commonly referred to as Parameters-based PTFs (e.g., Børgesen et al., 2006; Weber et al., 2024). This approach holds significant promise for enhancing PTF performance and has gained further attention (Ding & El-Zein, 2024). For example, Minasny and McBratney (2002) demonstrated a 13% reduction in root-mean-square-error for water content predictions using SPINN-based PTFs via Feedforward Neural Networks (FNNs), compared to Parameters-based PTFs.

Despite these advances, current PTFs have a critical limitation: nearly all are based on unimodal soil hydraulic models and thus cannot capture bimodal SHPs resulting from the effects of soil structure. Soil structure, which reflects the arrangement of soil pores, significantly influences SHPs (Chandrasekhar et al., 2018; Romero-Ruiz et al., 2018; Wang et al., 2023) and plays an important role in regulating hydrological processes such as infiltration, surface runoff, and groundwater recharge (e.g., Fatichi et al., 2020; Rabot et al., 2018; Vereecken et al., 2022). One of the most pronounced effects of soil structure on SHPs is the sharp decrease in hydraulic conductivity near saturation, caused by macropores, which leads to bimodal HCCs (Coppola, 2000; Wang et al., 2023). For instance, Wang et al. (2025) reported an approximately 100-fold average decrease in hydraulic conductivity as the matric potential decreased from 0 to −6 cm, based on tests conducted on 350 soil samples. Another effect is the presence of bimodal pore size distribution, resulting in bimodal SWRCs and HCCs (Costa & Cavalcante, 2021; Dexter et al., 2008; Durner, 1994; Hadas, 1987).

Previous attempts have been made to incorporate soil structure into PTF development. For example, Rawls et al. (1993) applied pore size scaling to estimate macropore K_s, Lin et al. (1999) introduced a detailed system of morphometric indices and demonstrated that these indices were more effective at predicting macro- and micropore flow compared to traditional factors like texture, Bulk Density (BD), or organic matter content. Pachepsky and Rawls (2003) and Hassan et al. (2022) used aggregate-size distributions as a structural indicator, improving predictions of water retention. Li et al. (2014) established an empirical relation between soil grain-size distribution and the parameters of bimodal SWRC. However, these models are still rudimentary and unable to predict or describe all the parameters related to bimodal SHPs.

Therefore, this study aims to develop new PTFs capable of estimating bimodal SHPs. We employ two bimodal soil hydraulic models: the third version of the Fredlund and Xing (1994)-Wang et al. (2018) model, developed by Wang et al. (2023) and termed as the FXW-M3 model, which captures bimodal HCCs but remains unimodal for SWRCs, and the bimodal Fredlund and Xing (1994)-Wang et al. (2018) model, developed by Wang et al. (2025) and termed as the B-FXW model, which captures both bimodal SWRCs and HCCs but at the cost of one additional free-fitted parameter. Both models account for capillarity, adsorption forces, and the influence of macropores, with the B-FXW model also addressing bimodal particle size distribution effects. For comparison, we also apply the unimodal Van Genuchten (1980)–Mualem (1976) (VGM) model, which accounts only for capillarity.

To build the PTFs, we employed two methods: the FNNs approach to develop SPINN-based PTFs, and a widely-used machine learning technique, Random Forest (RF) (Breiman, 2001), to develop Parameters-based PTFs. We also evaluated the effects of individual versus simultaneous optimization of SWRC and HCC in PTF development.

2. Materials and Methods

2.1. Data Sets

We utilized two publicly available data sets for developing PTFs. The first is the UNsaturated SOil hydraulic DAtabase (UNSODA) (Nemes et al., 2001), which contains measurements from soils worldwide, and has been widely applied in PTF development (Rudiyanto et al., 2021; Schaap et al., 2001; Wang, Zhou, et al., 2022; Zhang et al., 2018; Zhang & Schaap, 2017). The second is a recently released data set by Hohenbrink et al. (2023a,

Table 1
Basic Statistics of Soil Input Variables, Including Maximum (Max), Minimum (Min), Mean, Standard Deviation, and Median for Sand, Silt, and Clay Content (%), Bulk Density (BD, g/cm^3), and Organic Carbon (OC, % by Mass)

Variable	Min	Max	Mean	SD	Median
Sand	3	99.8	43.6	29.2	42.6
Silt	3	90	38.6	23.2	37.1
Clay	0	66.3	17.8	13.2	15.4
BD	0.37	1.89	1.329	0.296	1.39
OC	0.04	19.26	2.283	2.343	1.57

Note. A total of 438 samples were available for sand, silt, clay, and bulk density, while OC data were available for 297 samples.

2023b), which contains measurements from undisturbed, mostly central European soil samples, and has not yet been used in the development of PTFs.

To ensure that the data were sufficient to analyze bimodality, Zhang et al. (2022) recommended that at least three data pairs (including θ_s and K_s) are required for both SWRC and HCC when $h > -20$ cm. Only 47 soil samples from the two databases met this requirement. Therefore, we adjusted the criteria as follows: (a) measured K_s, soil texture information, and BD must be available; and (b) both $\theta(h)$ and $K(h)$ should include more than five measured data pairs; (c) for SWRC, at least two data pairs are required when $h > -6$ cm, a threshold above which macropores play a crucial role, as suggested by Jarvis (2020). As a result, a total of 438 soil samples were selected, 336 from the Hohenbrink data set and 104 from the UNSODA data set. On average, each soil sample in the Hohenbrink data set includes 194 θ–h pairs and 67 K–h pairs, whereas each sample in the UNSODA data set contains an average of 12 θ–h pairs and 18 K–h pairs.

Table 1 summarizes the basic statistics of the soil input variables for the selected soil samples, while Table 2 presents the statistical distribution of the measured SHPs across five matric potential ranges. In total, 66,184 data points were used for water retention and 24,348 for hydraulic conductivity. The majority of the water retention data (97.2%) correspond to matric potentials higher than −1,000 cm, with 33.6% above −20 cm. In contrast, approximately 91.8% of the measured hydraulic conductivity data fall within the matric potential range of −20 to −1,000 cm, while only 3.8% are at matric potentials above −20 cm.

Following Zhang et al. (2022), we used the Akaike Information Criterion (AIC) (see Section 2.5) to compare the fits of unimodal and bimodal VGM (Durner, 1994) models to determine whether the soil samples exhibited bimodal characteristics. Based on SWRC alone, 280 soil samples from the Hohenbrink data set and 21 samples from the UNSODA data set showed evidence of bimodality. When both SWRC and HCC were considered, 186 samples from the Hohenbrink data set and 52 samples from the UNSODA data set exhibited bimodal behavior.

To further validate the performance of the developed PTFs, we used 59 soil samples from Vereecken et al. (1989) (Table S1–S2 in Supporting Information S1). These samples meet the criteria set by Zhang et al. (2022) and are suitable for analyzing bimodality. According to the AIC value, 33 out of 59 soil samples exhibited bimodal behavior. Figure 1 presents the texture distribution of all soil samples used in the study.

2.2. The Tested Soil Hydraulic Models

We applied three soil hydraulic models for PTF development. A brief description of each model is provided below:

2.2.1. The VGM Model

The VGM model, a classic capillary-based, unimodal soil hydraulic model, is the most commonly used model for PTF development. The SWRC is expressed as:

Table 2
Basic Statistics of Soil Water Retention Curves and Hydraulic Conductivity Curves Across Five Ranges of Matric Potential (h, cm), Including the Mean, Standard Deviation, and Number of Observations for Volumetric Water Content (θ, cm^3/cm^3) and Hydraulic Conductivity (K, cm/d)

−h	Mean θ	SD of θ	Num of θ	Mean of K	SD of K	Num of K
0–6	0.49	0.10	10,273	1,454.40	3,226.83	535
6–20	0.46	0.10	11,967	18.90	112.10	398
20–1,000	0.35	0.11	42,072	0.31	10.78	22,344
1,000–15000	0.22	0.10	977	1.6e−3	2.57e−3	1,067
>15,000	0.05	0.05	895	1e−4	0.00	4

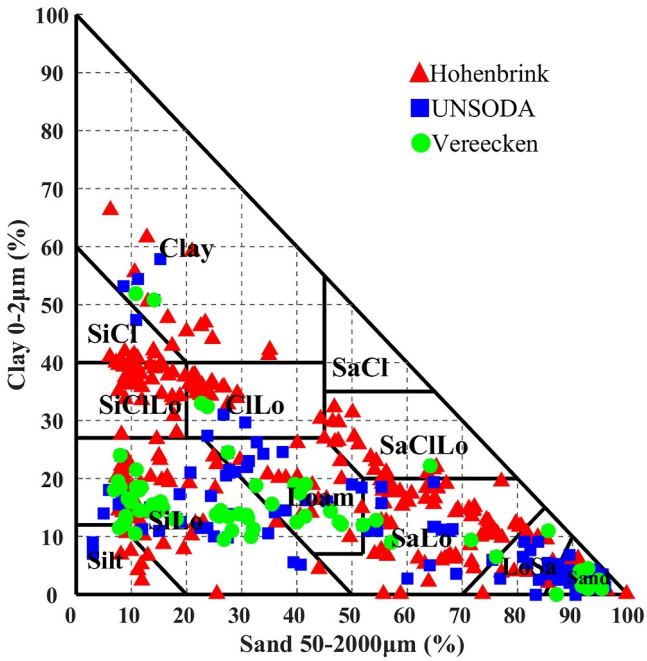

Figure 1. Soil texture distribution of the UNsaturated SOil hydraulic DAtabase, Hohenbrink and Vereecken data sets.

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} = [1 + (\alpha h)^n]^{-m} \tag{1}$$

where S_e is the effective water saturation degree related to capillary water. θ, θ_s, θ_r (cm^3/cm^3) are the volumetric, saturated and residual soil water contents, respectively, and α (cm^{-1}), n, and $m = 1 - 1/n$ are fitting parameters.

The HCC is written as:

$$K = K_s S_e^{\ l}\left(1 - \left(1 - S_e^{\ 1/m}\right)^m\right)^2 \tag{2}$$

where K and K_s (cm/day) are the hydraulic conductivity and saturated hydraulic conductivity, respectively, and l is set as 0.5.

Thus, the VGM model has five free-fitted parameters (α, n, θ_r, θ_s, K_s) in total to describe SHPs.

2.2.2. The FXW-M3 Model

The FXW-M3 model (Wang et al., 2023) considers the effects of capillarity, adsorption forces, and soil structure, particularly macropores. However, it does not account for the influence of bimodal particle size distribution, making it capable of describing bimodal HCCs but only unimodal SWRCs. Compared to the VGM model, the FXW-M3 model greatly improves SHP description at both the wet and dry ends (Wang et al., 2023). The SWRC of the FXW-M3 model comes from Fredlund and Xing (1994) and is written as:

$$\theta = \theta_s\left[1 - \frac{\ln(1 + h/h_r)}{\ln(1 + h_0/h_r)}\right]\Gamma(h) \tag{3}$$

where

$$\Gamma(h) = [\ln(e + |\alpha h|^n)]^{-m} \tag{4}$$

where h_0, setting as -6.3×10^6 cm, is the matric potential corresponding to zero water content (Schneider & Goss, 2012). h_r, setting as -1.5×10^4 cm, is a shape parameter (Fredlund & Xing, 1994). α (cm^{-1}), n and m are fitting parameters.

The HCC of the FXW-M3 model is divided into two segments representing the effects of soil structure and soil matrix, respectively. It is expressed as:

$$K(h) = \begin{cases} K(h_a)\left(\dfrac{\Gamma(h) - \Gamma(h_0)}{\Gamma(h_a) - \Gamma(h_0)}\right)^l \left[\dfrac{1 - \left(1 - \Gamma(h)^{1/m}\right)^{1-1/n}}{1 - \left(1 - \Gamma(h_a)^{1/m}\right)^{1-1/n}}\right]^2 & h < h_a \\ K_s \Gamma(h)^{\frac{\ln(K(h_a)/K_s)}{\ln(\Gamma(h_a))}} & h \geq h_a \end{cases} \tag{5}$$

where h_a is the critical matric potential separating soil structure and soil matrix effects, $K(h_a)$ (cm/day) is the saturated matrix hydraulic conductivity. Here, we treat h_a as a constant of -6 cm following Jarvis (2020) and $K(h_a)$ as a free-fitted parameter. Notably, in Wang et al. (2023), h_a is optimized and varies across different soil types. $K(h_a)$, theoretically estimated from the SWRC, exhibits high uncertainty, which leads to suboptimal

characterization of HCCs in certain soil samples. Overall, the FXW-M3 model applied here has six parameters (α, n, m, θ_s, $K(h_a)$, K_s).

2.2.3. The B-FXW Model

The B-FXW model (Wang et al., 2025) further considers the effects of bimodal particle size distribution, allowing it to describe both bimodal SWRCs and HCCs. Its form is similar to the FXW-M3 model, but with a different correction factor to capture bimodal behavior. The SWRC is expressed as:

$$\theta = \theta_s C(h)\Gamma(h) \tag{6}$$

where $\Gamma(h)$ is described by Equation 4 and the correction factor $C(h)$ is defined as

$$C(h) = \left[1 - \frac{\ln(1 + h/h_r)}{\ln(1 + h_0/h_r)}\right]^{n_c} \tag{7}$$

where n_c is a free-fitted parameter. Notably, when $n_c = 1$, Equation 6 reduces to Equation 3.

The HCC is described as

$$K = K_c + K_f \tag{8}$$

where K_c represents the hydraulic conductivity accounting for both capillary forces and soil structure, expressed as:

$$K_c(h) = \begin{cases} K(h_a)\left(\frac{S(h)}{S(h_a)}\right)^{l}\left[\frac{F(h)}{F(h_a)}\right]^2 & h < h_a \\ K_s S(h)^{\frac{\ln(K(h_a)/K_s)}{\ln(S(h_a))}} & h \geq h_a \end{cases} \tag{9}$$

where $S = \theta/\theta_s$ and $F(h)$ is expressed as:

$$\begin{aligned} \mathrm{F}(h) &= \int_{S(h_0)}^{S(h)} \frac{1}{h} dS = \int_{h_0}^{h} \frac{1}{h}\frac{dS}{dh} dh \\ &= \frac{1}{\theta_s}\int_{h_0}^{h} \frac{1}{h}(\theta_s \Gamma(h))\, \frac{dC(h)}{dh} dh + \frac{1}{\theta_s}\int_{h_0}^{h} \frac{1}{h}\theta_s C(h)\, \frac{d\Gamma(h)}{dh} dh \end{aligned} \tag{10}$$

Equation 10 needs to be solved numerically.

K_f in Equation 8 represents the hydraulic conductivity due to adsorption forces and is given by Wang, Ma, and Zhu (2022) as:

$$K_f(h) = b\theta_m(1 - \Gamma(h))\left(\frac{h}{h_m}\right)^{-1.5} \tag{11}$$

where h_m is the matric potential dominated by van der Waals forces, set at -1.0×10^5 cm following Lebeau and Konrad (2010), and θ_m is the corresponding water content, which can be estimated using Equation 6. The parameter b, set as 4.062×10^{-6} cm/d following Wang, Zhou, et al. (2022), represents the combined effects of film thickness, specific surface area, and a correction factor on the estimation of hydraulic conductivity. A detailed description of the B-FXW model can be found in Wang et al. (2025). In total, the B-FXW model applied here has seven free-fitted parameters ($\alpha, n, m, \theta_s, n_c, K(h_a), K_s$). Figure 2 illustrates the SWRC and HCC curves for the three soil hydraulic models.

The three models were fitted to the data. First, the SWRC was optimized using observed data to derive the corresponding soil hydraulic parameters, P_θ. The objective function $\Phi(P_\theta)$, to be minimized, is defined as:

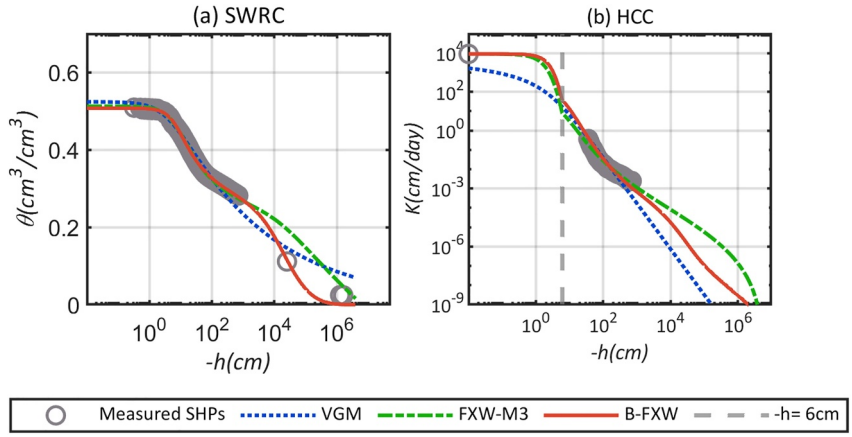

Figure 2. Illustration of the VGM, FXW-M3 and B-FXW models in describing (a) the SWRC and (b) the hydraulic conductivity curvel. Measured soil hydraulic properties are from sample KIT116 in the Hohenbrink data set. All three models were fitted using the shuffled complex evolution optimization algorithm.

$$\Phi(P_\theta) = \frac{\sum_{i=1}^{n_\theta} \left[\theta_i - \overline{\theta}_i(P_\theta)\right]^2}{n_\theta} \tag{12}$$

where θ_i and $\overline{\theta}_i(P_\theta)$ are the measured and simulated soil water contents, respectively, and n_θ is the number of data pairs. For the VGM model, $P_\theta = [\alpha, n, \theta_r, \theta_s]$; for the FXW-M3 model, $P_\theta = [\alpha, n, m, \theta_s]$; and for the B-FXW model, $P_\theta = [\alpha, n, m, \theta_s, n_c]$. Table 3 summarizes the fitted parameters for all three hydraulic models.

Next, the HCC was optimized using observed hydraulic conductivity data to derive parameters P_K. The objective function $\Phi(P_K)$ is defined as:

$$\Phi(P_K) = \frac{\sum_{j=1}^{n_K} \left[\log_{10}(K_j) - \log_{10}\left(\overline{K}_j(P_K)\right)\right]^2}{n_k} \tag{13}$$

where K_j and $\overline{K}_j(P_K)$ represent the measured and simulated hydraulic conductivities, respectively, and n_K is the number of data pairs. For the VGM model, $P_K = [K_s]$; for the FXW-M3 and B-FXW models, $P_K = [K(h_a), K_s]$.

Both Equations 12 and 13 were optimized using a global optimization method known as shuffled complex evolution, developed by Duan et al. (1992), implemented through the Python enviroment.

Table 3
The Upper and Lower Bounds of the Optimized Parameters for the Three Hydraulic Models

Model	Parameter	Lower boundary	Upper boundary
VGM	α (cm^{-1})	0.001	0.25
	n	1	15
	θ_r (cm^3/cm^3)	0	0.25
	θ_s (cm^3/cm^3)	0.2	0.85
	K_s (cm/d)	10^{-4}	$10^{4.5}$
FXW-M3/B-FXW	α (cm^{-1})	0.001	0.2
	n	1.01	15
	m	0.01	1.5
	θ_s (cm^3/cm^3)	0.2	0.85
	K_s (cm/d)	10^{-4}	$10^{4.5}$
	$K(h_a)$ (cm/d)	10^{-4}	10^{4}
	n_c	1	15

Note. Notably, n_c is only optimized for the B-FXW model.

2.3. PTFs Development

For each soil hydraulic model, we developed five PTFs which were designated as H1 to H5. Among these PTFs, H1 uses soil texture class as input, H2 incorporates sand, silt, and clay percentages, H3 adds the additional input of BD, H4 includes BD and the detailed particle size fractions in seven size ranges: <2 μm, 2–6.3 μm, 6.3–20 μm, 20–63 μm, 63–200 μm, 200–630 μm and 630–2000 μm, and H5 further includes Organic Carbon (OC) content. Notably, only 297 soil samples from Hohenbrink et al. (2023a, 2023b) met the requirement for building H4 and H5 PTFs.

To reduce bias associated with relying solely on a single train/test split and to ensure the robustness and generalizability of model performance, a K-fold cross-validation strategy ($K = 5$) was adopted, following Rodriguez et al. (2010). First, the soil samples were evenly divided into five subsets. In

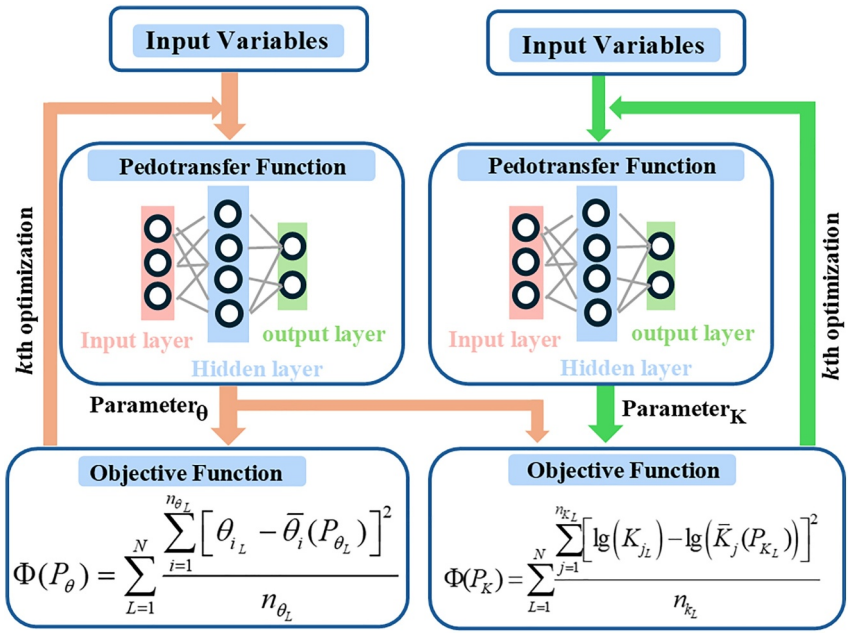

Figure 3. Process for developing the SPINN-based Pedotransfer functions (PTFs). The Feedforward Neural Network (FNN) on the left is initially trained to estimate the hydraulic parameters of the SWRC, which are then used as input for the FNN on the right to construct the PTF for estimating the parameters of the Hydraulic Conductivity Curves.

each iteration, one subset was used as the test set, while the remaining four subsets were used for training. This process was repeated five times, ensuring that each subset served as the test set exactly once. The model performance was then evaluated by averaging the results across all five folds, with the combined test sets covering the entire data set.

2.3.1. H1 PTF

The H1 PTF is a look-up table-based PTF. We classified the soil samples into 12 types based on USDA soil texture class. For each texture class, the hydraulic parameters of all corresponding soil samples were optimized and averaged to obtain representative values. The SWRC and HCC were optimized separately (Equations 12 and 13).

2.3.2. SPINN-Based PTFs Development

We applied the FNNs to develop SPINN-based PTFs (H2, H3, H4 and H5) with varying inputs. The FNN is a basic form of artificial neural network where information flows in one direction—from the input layer through hidden layers to the output layer—without loops or feedback (Bebis & Georgiopoulos, 1994; Shekhar et al., 1992). The number of hidden layers and hidden units was determined using Bayesian optimization (Wu et al., 2019) and was set to 2, 3 and 3 layers with 156, 374 and 226 hidden units for the VGM, FXW-M3 and B-FXW models, respectively. We employed the ReLu activation function to avoid gradient vanishing and used the Adam optimizer to adjust the network parameters and minimize the objective function.

SPINN predicts parameters of the hydraulic models, but the objective function to be minimized is the observed SWRC and HCC. In this way, the hydraulic model was embedded into the neural networks model (Rudiyanto et al., 2021).

Figure 3 illustrates the procedure for building the SPINN-based PTFs. In this process, the SWRC and HCC were optimized separately. First, we optimized the SWRC using all observed water retention data pairs in the training set with the FNN method to derive the PTF for estimating parameters P_θ. The estimated P_θ were then fed to another FNN to train the PTF for estimating P_K. The optimization functions, as shown in Figure 3, are based on observed water content and hydraulic conductivity rather than fitted soil hydraulic parameters, as used in

Parameters-based PTFs. The FNN optimization continued until the objective function stabilized and reached its minimum value.

Simultaneous optimization of SWRCs and HCCs has been widely used in PTFs development (e.g., Wang, Zhou, et al., 2022; Weber et al., 2020). To evaluate the impact of different optimization strategies on PTF performance, we also developed SPINN-based H3-J PTFs, where the objective function combines both water content and hydraulic conductivity, expressed as:

$$\Phi(P) = \sum_{L=1}^{N} \frac{\sum_{i=1}^{n_{\theta_L}} \left[\theta_i - \overline{\theta}_i(P_L)\right]^2}{n_{\theta_L} \times u_L} + \sum_{L=1}^{N} \frac{\sum_{j=1}^{n_{K_L}} \left[\log_{10}(K_j) - \log_{10}\left(\overline{K}_j(P_L)\right)\right]^2}{n_{k_L} \times v_L} \tag{14}$$

where $P = [P_\theta + P_K]$ represents the soil hydraulic parameters. N is the number of soil samples in the training set. u_L and v_L are weighting terms that balance the different magnitudes of $\theta(h)$ and $K(h)$. u_L is the mean square error for fitting SWRCs individually, and v_L is the mean square error for fitting HCCs individually. (Zhang et al., 2022).

2.3.3. Parameters-Based H3 PTF Development

As a comparison, we also developed Parameters-based H3 PTFs using the RF algorithm (Breiman, 2001). The same input variables as for other PTFs were used to predict parameters of the hydraulic parameters derived in Section 2.2. RF is considered one of the most effective machine learning methods and is widely used for predicting SHPs (Boulesteix et al., 2012; P. K. Gupta & Maiti, 2023; S. Gupta et al., 2021; Szabó et al., 2019). The RF method involves bagging, where samples are repeatedly drawn with replacement from the data set to create new training sets. Each training set generates a decision tree, and the diversity of these trees improves the model's predictive performance. In this study, we set the number of trees to 100 and the minimum number of leaves per tree to 10.

2.4. Comparison With Existing PTFs

We also included three existing PTFs developed by Zhang and Schaap (2017), Weynants et al. (2009), and Wang, Zhou, et al. (2022) for comparison, evaluated with the independent 59 soil samples from Vereecken et al. (1989). The PTFs from Zhang and Schaap (2017) and Weynants et al. (2009) are based on the VGM model, which accounts only for capillarity effects and is referred to here as the Rosetta3-PTF and Weynants-PTF, respectively. The PTF from Wang, Zhou, et al. (2022), termed the Wang-PTF, is based on the FXW-M1 model (Wang, Ma, & Zhu, 2022), which accounts for both capillarity and adsorption forces but does not account for soil structure. Table 4 lists the input variables, the applied methods and the predicted parameters for all PTFs.

2.5. Model Performance Metrics

The performance of the PTFs was evaluated using root mean square error (RMSE), mean error (ME), the adjusted coefficient of determination (R_a^2), and AIC. Both R_a^2 and AIC accounts for the impact of the number of free-fitted parameters.

The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(o_i - \overline{o}_i)^2} \tag{15}$$

The ME is defined as:

$$\text{ME} = \frac{\sum_{i=1}^{N}(\overline{o}_i - o_i)}{N} \tag{16}$$

The R_a^2 accounts for the number of independent variables in a model, eliminating their effect on the standard coefficient of determination (R^2). It is defined as:

Table 4
Detailed Information on All Pedotransfer Functions Used in This Study

PTFs	Reference	Methods for PTF development	Input variables	Predicted parameters
H1	This study	Average values of fitted parameters (obtained separately) for each soil texture	Soil Texture	VGM: $\alpha, n, \theta_r, \theta_s, K_s$. FXW-M3: $\alpha, n, m, \theta_s, K_s, K(h_a)$. B-FXW, $\alpha, n, m, \theta_s, n_c, K_s, K(h_a)$
SPINN-based H2		Separate optimization of SWRC and HCC using SPINN	Sand, Silt, Clay	
SPINN-based H3		Separate optimization of SWRC and HCC using SPINN	Sand, Silt, Clay, BD	
SPINN-based H4		Separate optimization of SWRC and HCC using SPINN	Particle size fractions, BD	
SPINN-based H5		Separate optimization of SWRC and HCC using SPINN	Particle size fractions, BD, OC	
SPINN-based H3-J		Simultaneous optimization of SWRC and HCC using SPINN	Sand, Silt, Clay, BD	
Parameters-based H3		Separate optimization of SWRC and HCC to derive parameters and develop PTF using Random Forest	Sand, Silt, Clay, BD	
Wang-PTF	Wang, Zhou, et al. (2022)	Simultaneous optimization of SWRC and HCC to derive parameters and develop PTF using Random Forest	Sand, Silt, Clay, BD	$\alpha, n, m, \theta_s, K_s$
Rosetta3-PTF	Zhang and Schaap (2017)	Optimization of SWRC to derive parameters and develop PTF using artificial neural network	Sand, Silt, Clay, BD	$\alpha, n, \theta_r, \theta_s, K_s$
Weynants-PTF	Weynants et al. (2009)	Empirical relationships between parameters and inputs are set a priori, with simultaneous optimization of SWRC and HCC used to determine the relationship parameters (PTF).	Sand, Silt, Clay, BD, OC	$\alpha, n, \theta_r, \theta_s, K_s$

$$R_a^2 = 1 - \frac{(1 - R^2)(N - 1)}{(N - p_m - 1)} \tag{17}$$

where R^2 is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(o_i - \overline{o}_i)^2}{\sum_{i=1}^{N}(o_i - \hat{o})^2} \tag{18}$$

AIC is based on the concept of entropy and quantifies both the complexity of a model and the accuracy of its fit to the data (Akaike, 1974). A lower AIC value indicates a model with a smaller prediction error and better overall performance. It is defined as

$$\mathrm{AIC} = N \times \ln(\mathrm{SSR}/N) + 2 \times p_m + \frac{2p_m(p_m + 1)}{N - p_m - 1} \tag{19}$$

where SSR is the sum of squared residuals, defined as:

$$\mathrm{SSR} = \sum_{i=1}^{n}(o_i - \overline{o}_i)^2 \tag{20}$$

For Equations 15–20, N is the number of data pairs, p_m is the number of model's parameters, and o_i and $\overline{o}_i$ are the measured and predicted values, respectively. $\hat{o}$ is the mean of o_i. Note that a log10 scale is applied to the K values (K in cm/d).

3. Results

3.1. Performance of Three Soil Hydraulic Models

Figure 4 summarizes the performance of the three soil hydraulic models when optimizing SWRCs and HCCs separately. All the aforementioned symbols are also summarized in Table A1. For SWRC fitting, the VGM model showed a significant overestimation for soil water content less than 0.1 cm^3/cm^3, as indicated by a high *ME* value

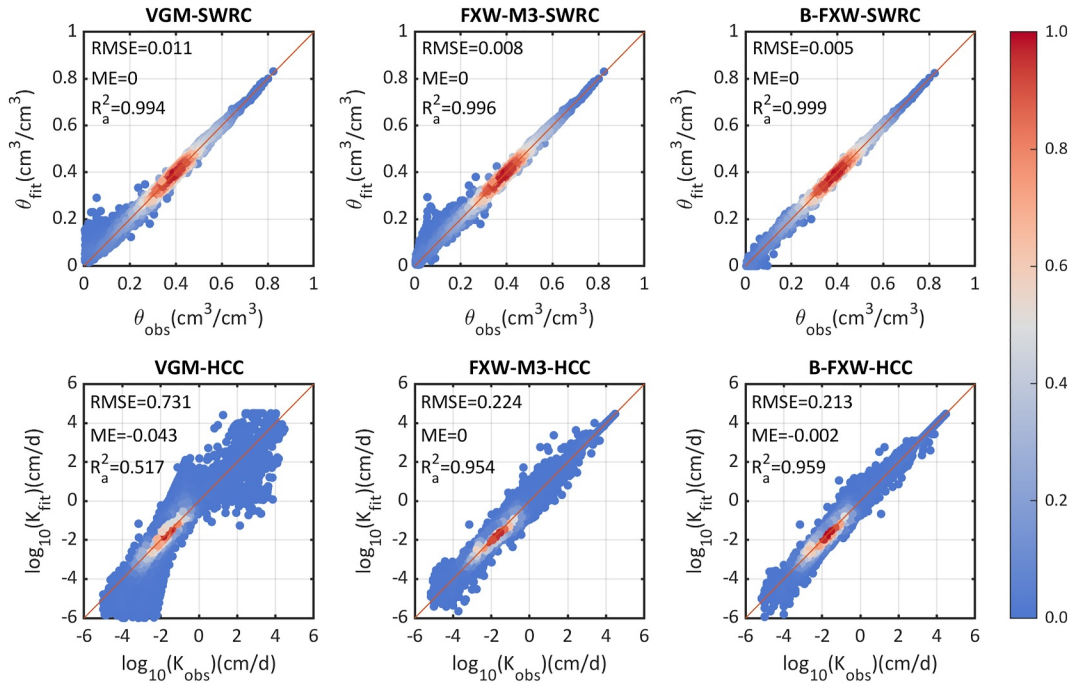

Figure 4. Optimized results of the three soil hydraulic models. The colorbar represents scatter densities, with red indicating high density and blue indicating low density. The solid red line denotes the 1:1 line. The subscript "obs" refers to the observed soil hydraulic properties, while the subscript "fit" represents the model-fitted SHPs.

of 0.020 cm^3/cm^3 within this moisture range. Consequently, the VGM model reported the highest $RMSE_\theta$ of 0.011 cm^3/cm^3 and the lowest R_a^2 of 0.994. The FXW-M3 model, which accounts for both capillary and adsorption forces, effectively reduced the *ME* value to 0.011 cm^3/cm^3 for soil water content below 0.1 cm^3/cm^3. The overall $RMSE_\theta$ is 0.008 cm^3/cm^3 and R_a^2 is 0.996. In contrast, the B-FXW model performed exceptionally well across the entire water content range, reporting the lowest $RMSE_\theta$ of 0.005 cm^3/cm^3 and the highest R_a^2 of 0.999. For soil water content below 0.1 cm^3/cm^3, the ME was -0.004 cm^3/cm^3.

For HCC, the VGM model exhibited poor performance overall, with the highest $RMSE_{\log_{10}(K)}$ of 0.731 and the lowest R_a^2 of 0.517. The model underestimated hydraulic conductivities in both wet and dry ranges for many soil samples. In comparison, both the FXW-M3 and B-FXW models performed well in fitting HCCs, with $RMSE_{\log_{10}(K)}$ values of 0.224 and 0.213, and R_a^2 values of 0.954 and 0.959, respectively. These results highlight the significant influence of soil structure in the selected samples.

3.2. PTFs Performances

3.2.1. Overall Performances of Soil Physics-Informed Neural Networks-Based PTFs

Figure 5 shows the results of the developed H1–H5 PTFs, which are based on three different soil hydraulic models in predicting SHPs. In general, with increased input of soil information, the developed PTFs demonstrated considerable improvements in performance from H1 to H3 PTFs (Figure 5 and Figures S1 and S2 in Supporting Information S1), aligning with findings from other research studies, such as Zhang and Schaap (2017) and Wang, Zhou, et al. (2022). For the three soil hydraulic models, the PTFs developed with the FXW-M3 model performed the best, followed by the B-FXW model and then by the VGM model. The results for the H1 and H2 PTFs are provided in the (Figures S1–S2, Tables S3–S7 in Supporting Information S1).

For the SPINN-based H4 and H5 PTFs, only 296 soil samples from the Hohenbrink et al. (2023a, 2023b) data set meet the data requirement. For the H4 PTF, the inputs included detailed particle size fractions across seven size ranges and BD, while the H5 PTF additionally required OC content.

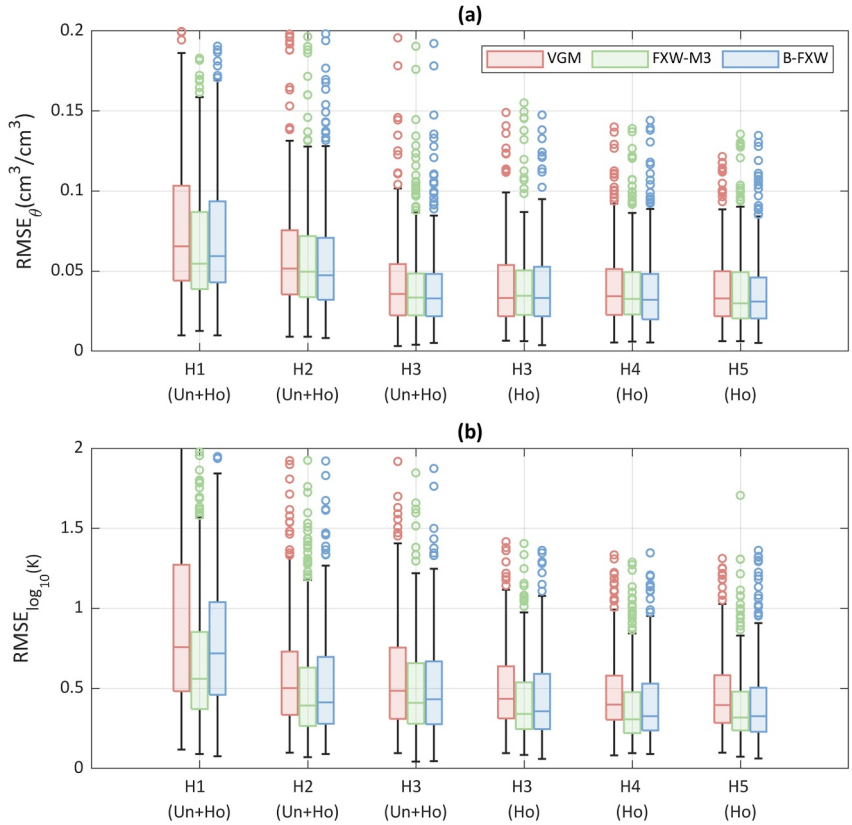

Figure 5. Performance of the developed H1–H5 Pedotransfer functions in predicting (a) SWRC and (b) Hydraulic Conductivity Curves. "Un + Ho" indicates that both the UNsaturated SOil hydraulic DAtabase (Nemes et al., 2001) Hohenbrink data sets (Hohenbrink et al., 2023a, 2023b) were used, whereas "Ho" indicates that only the Hohenbrink data set was applied. Data are aggregated from the test sets of 5-fold cross-validation.

From Figure 5, Figures S3 and S4 in Supporting Information S1, it is evident that including detailed particle size information (H4-PTFs) show negligible improvement in SWRC prediction in the test set. Both VGM and FXW-M3 model show almost the same $RMSE_\theta$ values for H3 and H4 PTFs while the mean $RMSE_\theta$ decreased slightly from 0.040 to 0.038 cm^3/cm^3 for the B-FXW model. In contrast, the H4-PTFs considerably enhanced the prediction of HCCs for all three models The reduction in $RMSE_{\log_{10}(K)}$ is 17.3%, 16.3% and 19.0% in the training set and 8.0%, 9.8% and 9.3% in the test set for VGM, FXW-M3 and B-FXW-based PTFs, respectively.

When OC content was added, the SPINN-based H5 PTFs showed a slight improvement in predicting SWRCs compared to the H4 PTFs for both training and test cases (Figure 5, Figures S4 to S5 and Tables S10-S11 in Supporting Information S1). For HCC prediction, the improvement was more pronounced for the VGM-based PTFs, with the $RMSE_{\log_{10}(K)}$ decreasing from 0.528 for the H4 PTF to 0.508 for the H5 PTF. In contrast, the PTFs developed with the FXW-M3 and B-FXW models showed almost no improvement in the test sets.

The results indicate that incorporating detailed particle size fractions has a relatively limited impact on improving SWRC prediction accuracy but contributes more significantly to enhancing HCC predictions. The further inclusion of OC content did not lead to substantial improvements in predictive performance. However, it should be noted that this analysis was based on only 296 soil samples, which may have constrained the assessment of its effectiveness.

3.2.2. Performance of H3 PTFs

For the applied data set, H3 PTFs had the most soil samples with the most input variables. Meanwhile, H3 PTFs are also the most applied in the literature (Børgesen et al., 2008; Rudiyanto et al., 2021; Selahvarzi et al., 2025;

Table 5
Performance of Three SPINN-Based H3 Pedotransfer Functions in Predicting Soil Water Retention Curves and Hydraulic Conductivity Curves in Test Sets

		SWRC-test			HCC-test		
	$-h$ (cm)	VGM	FXW-M3	B-FXW	VGM	FXW-M3	B-FXW
RMSE	0–20	0.04	0.039	0.039	1.101	0.836	0.925
	20–1,000	0.05	0.047	0.046	0.579	0.514	0.543
	>1,000	0.053	0.048	0.039	0.928	0.531	0.591
	ALL	0.047	0.044	0.043	0.626	0.531	0.564
R_a^2	0–20	0.834	0.844	0.842	0.499	0.71	0.645
	20–1,000	0.811	0.831	0.841	0.415	0.539	0.487
	>1,000	0.784	0.826	0.884	−0.602	0.476	0.349
	ALL	0.873	0.886	0.891	0.645	0.745	0.712
ME	0–20	−0.004	−0.003	0.001	−0.228	0.112	0.187
	20–1,000	−0.003	0.001	−0.001	0.157	0.099	0.033
	>1,000	0.012	0.022	0.001	−0.644	−0.072	−0.213
	ALL	−0.003	0.001	0.000	0.107	0.092	0.028

Note. The results represent the average from 5-fold cross validation. That log10 scale is applied to the K value (K in cm/d).

Wang, Zhou, et al., 2022; Zhang & Schaap, 2017). So, here and hereafter, we focus primarily on the performance of the H3 PTFs.

Table 5 presents the performance of the three SPINN-based H3 PTFs in predicting SWRC and HCC for test sets. The performance values are the average results from 5-fold cross-validation.

For SWRC prediction, the SPINN-based H3 PTFs developed with the VGM, FXW-M3, and B-FXW models demonstrated overall improved performance in test sets, as indicated by reduced $RMSE_\theta$ values, increased R_a^2, and ME values closer to zero. The overall improvement is more pronounced for the PTF developed with the FXW-M3 model compared to that with the VGM model. The B-FXW-based PTF showed only a slight enhancement over the FXW-M3-based PTF in terms of overall performance. However, when examining performance across different matric potential ranges, the SPINN-based H3-PTF developed with the B-FXW model significantly improved the prediction of soil water content for matric potentials lower than −1,000 cm. It achieved the lowest $RMSE_\theta$ (0.033 cm^3/cm^3), the highest R_a^2 (0.915), and an *ME* value of 0.002 cm^3/cm^3—closest to zero. In contrast, the SPINN-based H3 PTFs developed with the VGM and FXW-M3 models overestimated soil water content in this range, with higher *ME* values of 0.013 and 0.023 cm^3/cm^3, respectively.

In terms of predicting HCCs, the SPINN-based H3 PTF developed with the VGM model significantly underestimated hydraulic conductivity in both wet and dry ranges (Figure 6), with ME values of −0.228 and −0.644 in the test set for matric potential higher than −20 cm and less than −1,000 cm, respectively (Table 5). It also overestimated hydraulic conductivity in moderate matric potentials, as indicated by an ME value of 0.157 in the test set. Consequently, this PTF reported the highest $RMSE_{\log_{10}(K)}$ values of 0.596 and 0.626, and the lowest R_a^2 values of 0.678 and 0.645 in the training and test sets, respectively. The H3 PTFs developed with the FXW-M3 and B-FXW models demonstrated close agreement with observations across all measured matric potential ranges, significantly outperforming the VGM-based PTF, particularly for matric potentials greater than −20 cm and less than −1,000 cm. The SPINN-based H3 PTF developed with the FXW-M3 model performed the best, reporting the lowest $RMSE_{\log_{10}(K)}$ values of 0.474 and 0.531 in the training and test sets, respectively. The H3-PTF developed with the B-FXW model showed similar performance, with slightly higher $RMSE_{\log_{10}(K)}$ values of 0.509 and 0.564 in the training and test sets, respectively.

According to the *AIC* values, the SPINN-based H3 PTF developed with the FXW-M3 model outperformed the PTF developed with the VGM model for 277 soil samples and outperformed the B-FXW model for 245 soil samples (Data Set S1).

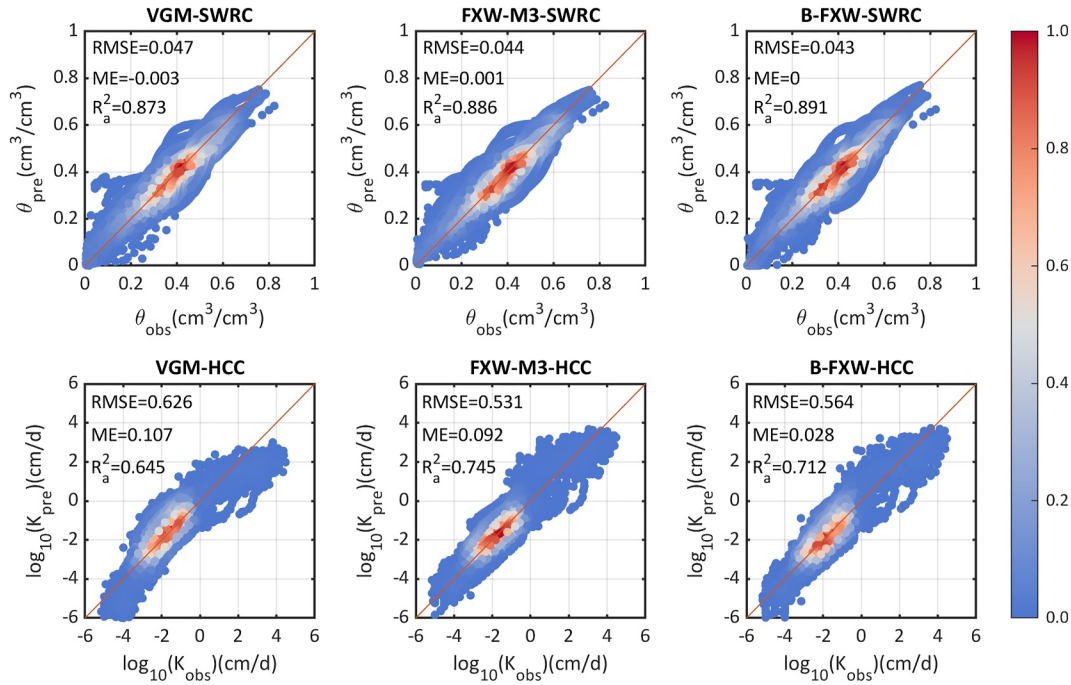

Figure 6. Prediction of soil hydraulic properties using SPINN-based H3 Pedotransfer functions developed with three soil hydraulic models. The inputs include the percentages of sand, silt, and clay, as well as bulk density. Data are aggregated from the test sets of 5-fold cross-validation.

To assess the ability of the developed PTFs to capture bimodal SHPs, we selected eight soil samples with varying soil types from the test data set, including four from the Hohenbrink et al. (2023a, 2023b) data set and four from the UNSODA data set, as shown in Figure 7. The SPINN-based H3 PTFs were applied to predict SHPs. For SWRC prediction, the B-FXW model demonstrated the highest flexibility and overall better performance. It successfully captured the bimodal SWRCs for soil samples KIT-130 (silty clay loam) and 2652 (loam), but failed to do so for soil sample 4091 (silt loam).

A detailed examination of individual soil samples showed that the B-FXW model generally captured bimodal SWRCs in the training set but struggled to do so in the test set. In contrast, both the FXW-M3 and VGM models were unable to capture bimodal SWRCs and generally overestimated the water contents in the dry range, particularly for soil samples from the Hohenbrink et al. (2023a, 2023b) data set.

For HCC prediction, both the FXW-M3 and B-FXW models successfully captured the bimodal shape for most soil samples, whereas the VGM model significantly underestimated hydraulic conductivity for both near saturation (KIT-109, KIT-130, KIT-113 and 4091) and dry conditions (KIT-130, TUBS-206 and 4651).

3.3. Comparison With Classic PTFs Using an Independent Data Set

Figure 8 shows the performance of the six PTFs in predicting SHPs, with the red dots representing data where the matric potential is higher than −20 cm, a threshold suggested by Zhang et al. (2022) to separate the effects of soil structure and soil texture. For SWRC prediction, the performance of the six PTFs was similar. The SPINN-based H3-PTFs developed here with the FXW-M3 and B-FXW models reported an $RMSE_\theta$ value of 0.038 cm^3/cm^3. The Wang-PTF and Rosetta3-PTF performed slightly better, with an $RMSE_\theta$ value of 0.037 cm^3/cm^3, while the Weynants-PTF had the lowest $RMSE_\theta$ value of 0.034 cm^3/cm^3. The SPINN-based H3-PTFs developed here with the VGM model performed the worst, with an $RMSE_\theta$ value of 0.040 cm^3/cm^3. The superior performance of the Weynants-PTF is unsurprising, as it was calibrated using the Vereecken et al. (1989) data set. Additionally, it should be noted that the measured water retention data do not decrease monotonically with matric potential (Figure 9), which greatly affects $RMSE_\theta$ values.

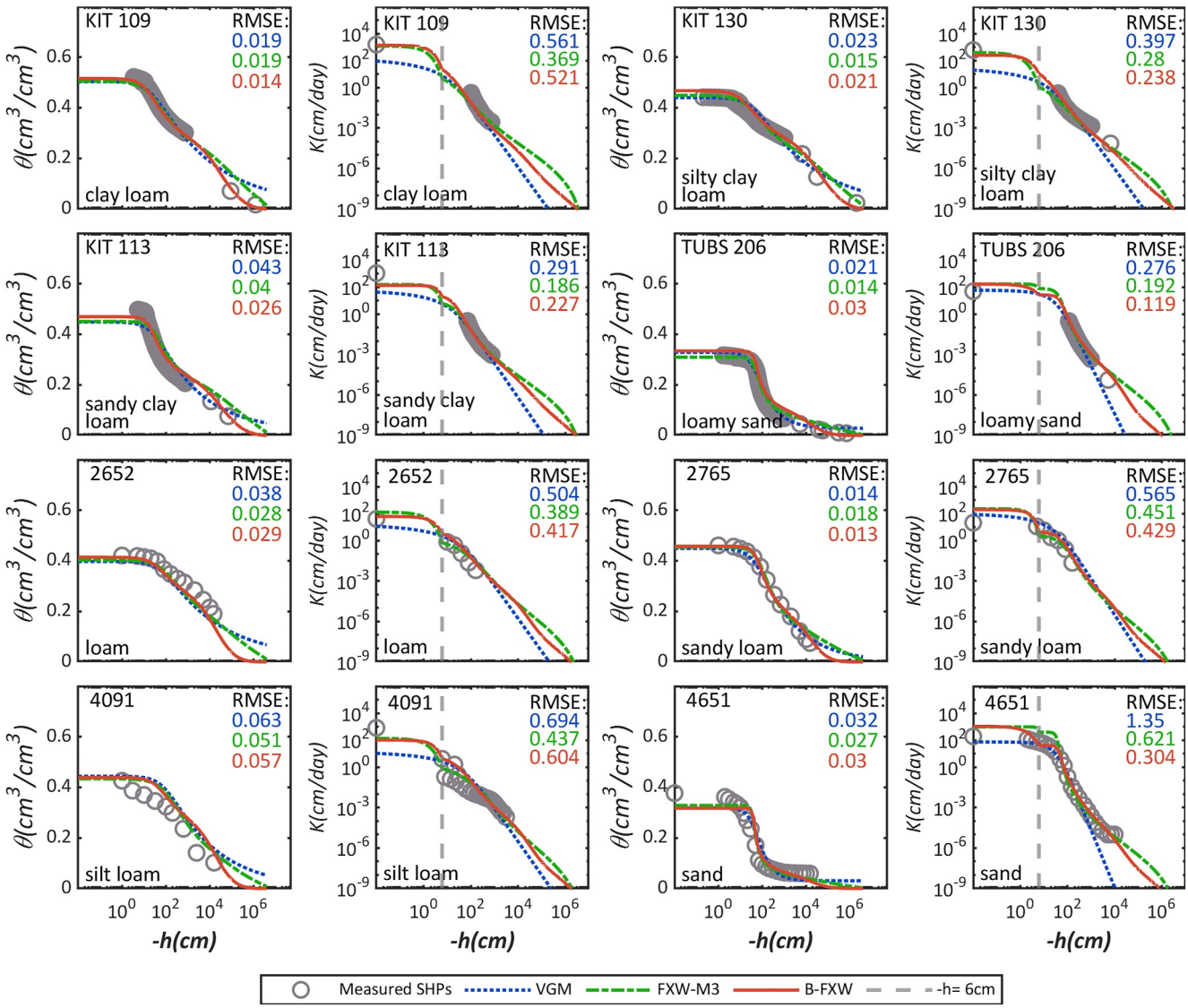

Figure 7. Predictions of soil hydraulic properties with three SPINN-based H3 Pedotransfer functions for eight soil samples in the test set.

For HCC prediction, the performance of the six PTFs varied considerably. The SPINN-based H3 PTF developed here with the FXW-M3 model performed the best, followed by the Wang-PTF, and then the PTF developed here with the B-FXW model, reporting $RMSE_{\log 10(K)}$ values of 0.635, 0.674, 0.842, respectively. Although the Wang-PTF performed well overall, it underestimated hydraulic conductivities for many soil samples where the matric potential was higher than—20 cm (Figure 8). In contrast, the PTFs developed with the FXW-M3 and B-FXW models improved performance near saturation. The three VGM model-based PTFs performed much worse, with $RMSE_{\log 10(K)}$ values of 1.275, 1.626, and 1.71 for the PTF developed here, the Weynants-PTF and the Rosetta3-PTF, respectively. It was clear that the unimodal VGM model-based PTFs underestimated hydraulic conductivity under both near-saturation and dry conditions.

According to the *AIC* values, when considering both SWRC and HCC predictions, the SPINN-based H3-PTFs developed with the FXW-M3 model performed best for 32 out of 59 soil samples, followed by the Wang-PTF, which performed best for 20 samples (Data Set S1).

The performance of the PTFs with eight selected soil samples, as shown in Figure 9, clearly demonstrated that the PTFs developed with the FXW-M3 and B-FXW models effectively captured the bimodal HCCs near saturation. While the Rosetta3-PTF also predicted K_s well, it overestimated unsaturated hydraulic conductivity in the mid-

10.1029/2024WR039337

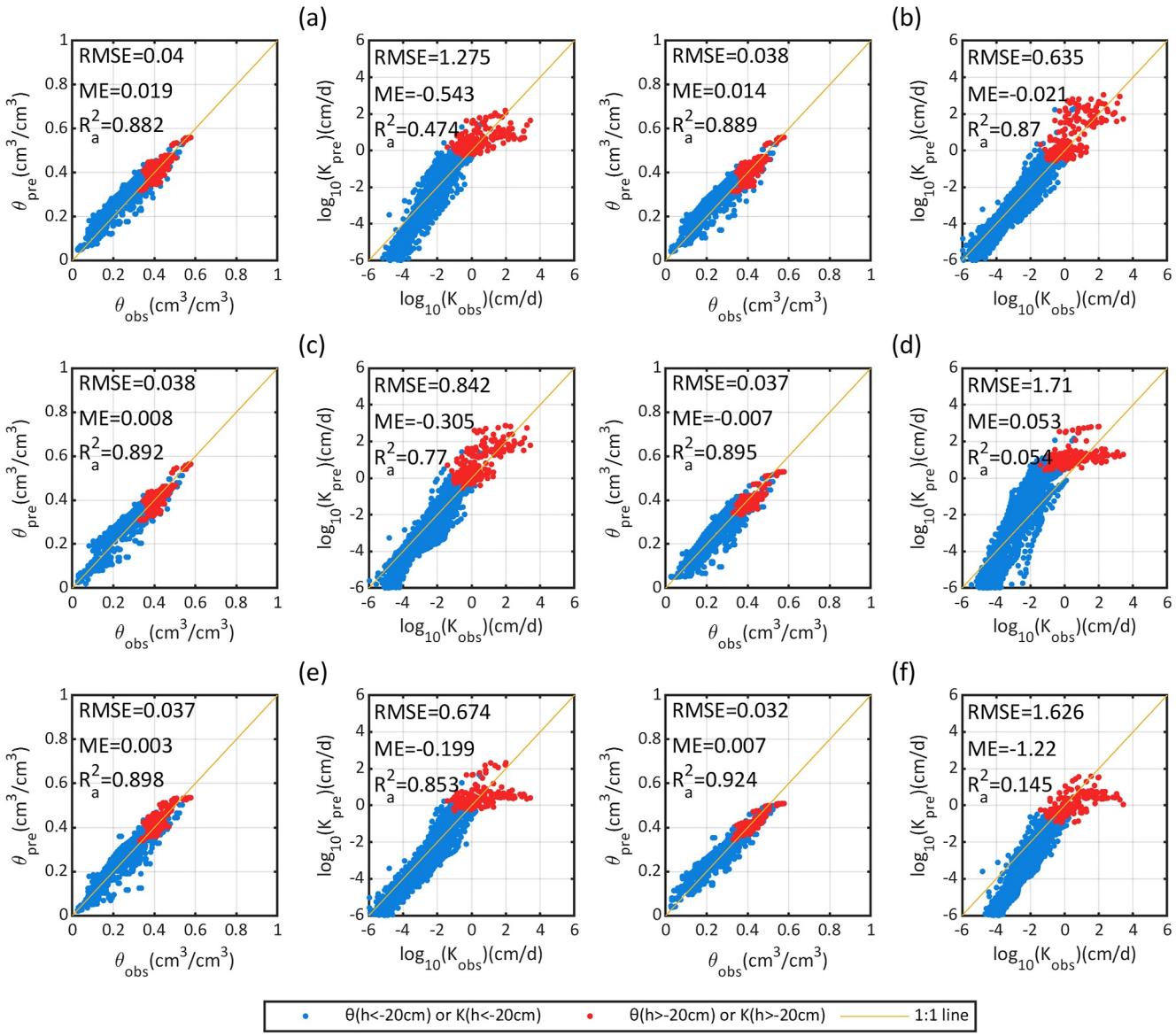

Figure 8. Prediction of soil hydraulic properties for 59 soil samples from Vereecken's data set using six Pedotransfer functions (PTFs). (a) SPINN-based PTF developed with the VGM model in this study; (b) SPINN-based PTF developed with the FXW-M3 model in this study; (c) SPINN-based PTF developed with the B-FXW model in this study; (d) Rosetta3 PTF that developed by Zhang and Schaap (2017); (e) Wang-PTF developed by Wang, Zhou, et al. (2022); (f) Weynants-PTF developed by Weynants et al. (2009).

saturation range and underestimated it in the dry range. In contrast, the Weynants-PTF performed better in the mid-saturation range but underestimated hydraulic conductivity both near saturation and in dry conditions. The difference in performance between the Rosetta3-PTF and Weynants-PTF is attributed to the fact that the Rosetta3-PTF used measured K_s for PTF development (Zhang & Schaap, 2017), whereas the Weynants-PTF used optimized K_s based on observed HCC for PTF development (Weynants et al., 2009).

4. Discussion

4.1. The Impact of Inputs on PTF Performance

In this study, we developed five types of PTFs with increasing amounts of input data. The results indicate that PTFs with more input variables generally perform better in predicting SHPs. The H2 PTFs, which used sand, silt, and clay percentages, significantly outperformed the H1-PTFs that used only soil class as input. The H3 PTFs,

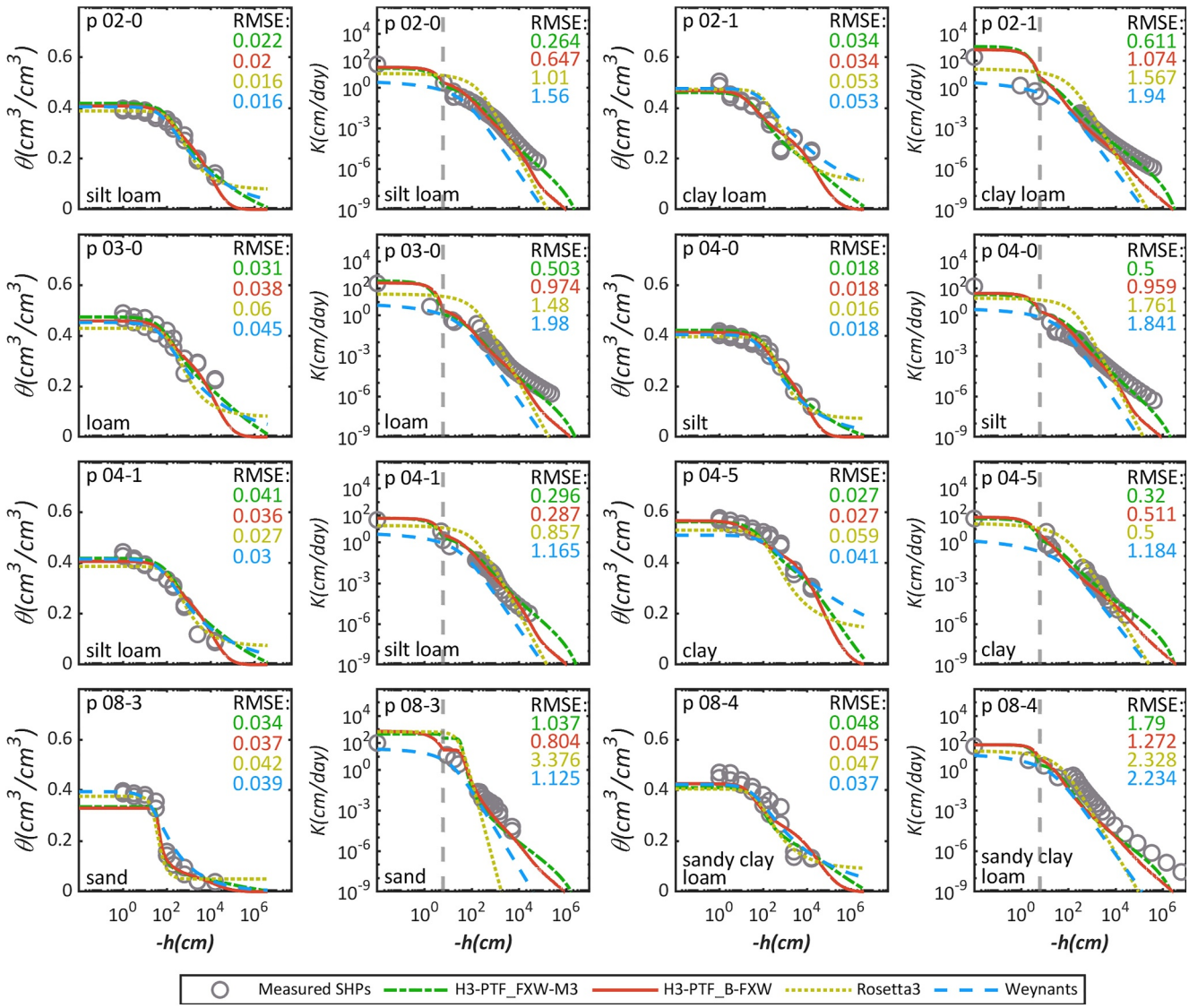

Figure 9. Predictions of soil hydraulic properties for eight soil samples from Vereecken's data set using four Pedotransfer functions.

which incorporated BD, showed substantial improvement over the H2 PTFs. These three PTFs are commonly found in the literature, and our findings align well with previous studies such as Wang, Zhou, et al. (2022).

For the 296 soil samples from the Hohenbrink et al. (2023a, 2023b) data set, which includes detailed particle size distribution across seven size ranges and OC content, we also developed H4 PTFs and H5 PTFs. The evaluation data showed that using detailed particle size information led to significant improvements in HCC predictions, while including OC content slightly improved SWRC predictions. While these types of PTFs are less frequently reported in the literature, our results confirm that, in general, the more information included in the inputs, the better the PTFs perform in predicting SHPs.

However, there remains a lack of representation of soil structural information in the applied inputs, as evidenced by the underestimation of soil water content near saturation and the relatively poor estimation of K_s for some soil samples across all developed PTFs (Figures 6–9).

4.2. The Impact of Soil Hydraulic Models on PTF Performance

We employed three soil hydraulic models—VGM, FXW-M3, and B-FXW—in PTFs development. These models showed an increasing ability to describe SHPs, reflected in their improved performance in fitting SHPs (Figure 4), though at the cost of increased model complexity, reflected by a higher number of free-fitted parameters (Table 3).

The results shown in Figure 5 through 9 indicate a strong influence of the soil hydraulic models on PTF performance. PTFs developed using the VGM model tended to overestimate water content and underestimate hydraulic conductivity under dry conditions for many soil samples. This overestimation and underestimation pattern has also been observed in other studies (Rudiyanto et al., 2021; Wang, Zhou, et al., 2022; Weber et al., 2020), and is primarily attributed to the VGM model's neglect of adsorption forces, which are dominant in dry conditions (Rudiyanto et al., 2021; Tuller & Or, 2001; Wang, Zhou, et al., 2022; Wang et al., 2016). The improved performance in dry conditions by the PTFs developed with the FXW-M3 and B-FXW models can be attributed to their representation of adsorption forces, as accounted for in these models (Wang et al., 2023, 2025).

The PTFs developed with the VGM model also significantly underestimated hydraulic conductivity at observed values above approximately 1 cm/day (Figures 6 and 8), a result also noted in previous studies (e.g., Manyame et al., 2007; Rudiyanto et al., 2021; Wang, Zhou, et al., 2022). This is because the VGM model does not account for soil structure effects, particularly the presence of macropores, which are known to lead to bimodal HCCs occurring near saturation (Børgesen et al., 2006; Mohanty et al., 1997; Wang et al., 2023, 2025). In contrast, both the FXW-M3 and B-FXW models account for the effects of soil structure due to macropores (Wang et al., 2023, 2025), enabling the PTFs developed with these models to capture bimodal HCCs and significantly improve HCC predictions near saturation (Figures 6–9). The presence of macropores and their impact on HCCs is commonly observed in nature (Coppola, 2000; Dexter et al., 2008; Hadas, 1987) and plays a crucial role in soil water flow and related processes (Beven & Germann, 1982; Gerke & van Genuchten, 1993). Therefore, the PTFs developed here with the FXW-M3 and B-FXW models provide more reliable SHPs estimates for field applications.

However, it should be noted that the PTFs developed with the FXW-M3 and B-FXW models still underestimated hydraulic conductivity at the wet end for some soil samples (Figure 7), likely due to limited soil structure information in the input variables. The saturated hydraulic conductivity K_s is heavily influenced by soil structure, particularly macropores (Bottinelli et al., 2016; Muller, 2018; Wang et al., 2023). Zhang et al. (2023) indicated that macropore volume and density are positively correlated with K_s, but accurately estimating K_s using only soil texture and BD remains challenging (Jorda et al., 2015; Toth et al., 2015; Zhang et al., 2018; Zhang & Schaap, 2019).

Compared to the FXW-M3 model, the B-FXW model also accounts for soil structure effects due to bimodal particle size distribution, at the cost of one additional parameter. The B-FXW model, developed by Wang et al. (2025), captures both bimodal SWRCs and HCCs. Optimization results showed that the B-FXW model outperformed the FXW-M3 model in fitting both SWRCs and HCCs (Figure 4). However, while the PTF developed with the B-FXW model showed slight improvement in predicting SWRCs, it performed worse in predicting HCCs compared to the FXW-M3-based PTF (Figure 6). This may be due to the B-FXW model's seven free-fitted parameters, which increase uncertainty in PTF development, as well as the limited input variables that may not fully characterize soil properties (Qin et al., 2024). Additionally, PTF with the B-FXW model was often unable to capture bimodal SWRCs in the test set, likely due to the small number of bimodal SWRC samples in the data set and a lack of sufficient input information to reflect bimodal characteristics.

Based on the AIC values, the PTFs developed with the FXW-M3 model outperformed those developed with the other two models for most soil samples, making them the most recommended.

4.3. The Impact of Optimization Methods on PTFs Development

To examine the impact of different optimization strategies, we also developed SPINN-based H3-J PTFs using a single FNN model to optimize both SWRC and HCC simultaneously. As shown in Table 6, Table S12 and Figure S6 in Supporting Information S1, notable differences in model performance were observed between the two optimization methods. In general, H3-J PTFs developed through simultaneous optimization of SWRCs and HCCs showed worse performance in predicting SWRCs, but improved performance in predicting HCCs, compared to those developed using separate optimization. For the VGM model, the $RMSE_\theta$ increased from 0.047 cm^3/cm^3

Table 6
Root Mean Square Error Values of SPINN-Based H3, H3-J and Parameters-Based H3 Pedotransfer functions in Predicting Soil Water Retention Curves and Hydraulic Conductivity Curves in the Test Sets

	SWRC-test			HCC-test		
PTF	VGM	FXW-M3	B-FXW	VGM	FXW-M3	B-FXW
SPINN-based H3	0.047	0.044	0.043	0.626	0.531	0.564
SPINN-based H3-J	0.052	0.048	0.049	0.605	0.504	0.539
Parameters-based H3	0.060	0.051	0.051	0.922	0.559	0.588

Note. SPINN-based H3 and SPINN-based H3-J refer to the SPINN-based PTFs developed by separately and simultaneously optimizing SWRCs and HCCs, respectively. The parameter-based H3 PTF refers to the model developed using Random Forest, with separate optimization of the SWRCs and HCCs. that for HCC, log10 scale is applied to the K value (K in cm/d). The units of $RMSE_\theta$ is cm^3/cm^3.

(separate optimization) to 0.052 cm^3/cm^3, while $RMSE_{\log_{10}(K)}$ decreased from 0.626 to 0.605 in the test set. Similarly, the B-FXW model improved in HCC prediction under simultaneous optimization, with $RMSE_{\log_{10}(K)}$ decreasing from 0.564 to 0.539, but $RMSE_\theta$ increasing from 0.043 cm^3/cm^3 to 0.049 cm^3/cm^3. For FXW-M3 model, with $RMSE_{\log_{10}(K)}$ decreasing from 0.531 to 0.504, and $RMSE_\theta$ increasing from 0.044 cm^3/cm^3 to 0.048 cm^3/cm^3. Overall, the simultaneous optimization achieved an average 4.3% reduction in $RMSE_{\log_{10}(K)}$, while in the cost of an average 11.2% increase in $RMSE_\theta$.

The comparison between simultaneous and separate optimization of SWRCs and HCCs revealed that the simultaneous optimization approach improves HCC prediction performance but at the expense of SWRC prediction accuracy. This outcome is largely influenced by the weighting in the objective function. In the separate optimization approach, the objective function of the first FNN includes only SWRC, leading to parameter vector P_θ that focuses exclusively on SWRC, thus accurately predicting water content. However, in the simultaneous optimization method, P_θ must account for both SWRC and HCC, which reduces the priority on SWRC, improving hydraulic conductivity prediction while diminishing water content prediction performance. This trade-off is driven by the weighting of the objective function. Since the slope of the SWRC is crucial for simulating soil water flow dynamics, the results presented here provide no strong evidence to favor simultaneous optimization over separate optimization in developing PTFs.

4.4. Comparison Between SPINN-Based and Parameters-Based PTFs

To further illustrate the performance differences between Parameters-based and SPINN-based PTFs, Table 6, Table S13 and Figure S7 in Supporting Information S1 show the performance of the Parameters-based H3 PTFs. The results indicate that the performance of Parameters-based PTFs in predicting SWRC and HCC was inferior to that of SPINN-based PTFs. For the B-FXW model, the $RMSE_\theta$ increased from 0.043 cm^3/cm^3 for the SPINN -based PTF to 0.051 cm^3/cm^3 for the Parameters-based PTF, and the $RMSE_{\log_{10}(K)}$ increased from 0.564 to 0.588. For the FXW-M3 model, the $RMSE_\theta$ increased from 0.044 cm^3/cm^3 (SPINN-based PTF) to 0.051 cm^3/cm^3 (Parameters-based PTF), and $RMSE_{\log_{10}(K)}$ increased from 0.531 to 0.559. The VGM model showed the most significant change, with $RMSE_\theta$ increasing from 0.047 cm^3/cm^3 (SPINN-based PTF) to 0.060 cm^3/cm^3 (Parameters-based PTF) and $RMSE_{\log_{10}(K)}$ increasing from 0.626 to 0.922. When predicting water content, the $RMSE_\theta$ values were reduced here by 21.7%, 13.7%, and 15.7% for the VGM, FXW-M3, and B-FXW models, respectively. Minasny and McBratney (2002) demonstrated that using the VGM-PTF with SWRC as the objective function reduced $RMSE_\theta$ by 13.0%. For hydraulic conductivity prediction, the $RMSE_{\log_{10}(K)}$ values were reduced by 32.1%, 5.0%, and 4.1% for the VGM, FXW-M3, and B-FXW models, respectively. These results suggest that the SPINN-based method is more accurate than the traditional approach for developing PTFs.

The advantages of SPINN-based PTFs may be attributed to the procedure in PTFs development. That is, the development of Parameters-based PTFs requires two procedures: one is fitting the soil hydraulic model with observations to derive parameters, and the other is building PTFs between inputs and the fitted parameters by minimizing the errors in parameters estimation. Both processes involve uncertainty. In contrast, the SPINN-based

PTFs minimize errors of SHPs estimation directly, thus avoiding the uncertainties and errors associated with the parameter fitting process (Minasny et al., 2024).

4.5. Limitations on the PTFs

While this study successfully developed PTFs for three hydraulic models, certain limitations remain:

1. The input variables used in this study—percentages of sand, silt, clay, and BD—were insufficient to fully capture bimodal characteristics. Although the H5 PTFs, which included detailed particle size distribution, BD, and OC content, improved the prediction performance in the training set, they still failed to capture bimodal SWRCs in the test set. In fact, the training set still underestimated saturated hydraulic conductivity (K_s) near saturation, an issue that has also been noted in studies such as Araya and Ghezzehei (2019), Báťková et al. (2023), Schaap et al. (2001), and Wösten et al. (1998). All of these studies highlighted the challenge of accurately estimating K_s using only soil texture, BD, and organic matter content, as these inputs do not adequately represent soil structure, particularly macropores. Jena et al. (2021) suggested that incorporating additional parameters, such as uniformity coefficient and geometric mean diameter could enhance the ability of PTFs to predict K_s.
2. The data set used in this study included a limited number of soil samples and SHP observations. First, only 309 soil samples were utilized for PTF training, which is insufficient to account for the significant variation in soil properties encountered in the field. Second, there were limited observations in both near-saturation and dry conditions. For example, HCC measurements near saturation were scarce, and most samples had only a single observation, making it challenging to capture the effects of macropores. Additionally, for SWRC, there were too few soil samples with clearly defined bimodal characteristics, particularly in the Hohenbrink et al. (2023a, 2023b) data set. This is considered a primary reason why the PTF developed with the B-FXW model struggled to predict bimodal SWRCs effectively.

5. Conclusion

In this study, we developed new PTFs using bimodal soil hydraulic models and SPINN, addressing a key limitation of traditional PTFs, which are generally based on unimodal models and fail to capture the effects of soil structure. The results demonstrated that the new PTFs effectively captured the bimodal HCCs and significantly outperformed unimodal models, with reduced prediction errors and improved reliability for field applications. While the new PTFs also improved the estimation of SWRCs, challenges remain in accurately capturing bimodal SWRCs for certain soil samples, largely due to the limited number of samples with bimodal characteristics in the data set. Nevertheless, the results suggest that PTFs based on bimodal models provide more reliable estimates of SHPs, especially in capturing the influence of macropores. Specifically, the SHPs-based PTFs developed with the FXW-M3 model were most recommended when balancing model complexity with PTF performance.

A key finding was that directly optimizing measured SHPs using SPINN consistently outperformed the conventional approach of optimizing fitted hydraulic parameters via machine learning. Additionally, we found that the optimization strategy—whether optimizing SWRCs and HCCs separately or simultaneously—can significantly affect PTF performance. The average 4.3% reduction in $\mathrm{RMSE}_{\log_{10}(K)}$ while at the cost of 11.2% increase in RMSE_{θ}, suggested that separate optimization is more reasonable for developing SPINN-based PTFs.

Overall, the bimodal PTFs presented in this work offer a significant advancement in soil hydraulic property estimation. However, to enhance the applicability of PTFs, future studies could incorporate other soil hydraulic property databases that offer more extensive measurements and information in both wet and dry ranges for additional training and validation. Furthermore, the development of PTFs should incorporate additional input variables that better capture the effects of soil structure.

Appendix A

Table A1
All Symbols Applied in the Main Text

Symbols	Unit	Description
b	cm/d	represents the combined effects on the estimation of hydraulic conductivity
e	–	The Euler's number (the e constant)
h	cm	Matric suction
h_0	cm	matric potential corresponding to zero water content
h_a	cm	critical matric potential separating soil structure and soil matrix effects
h_m	cm	matric potential dominated by van der Waals forces
h_r	cm	matric suction corresponding to the residual water content
K	cm/d	the total hydraulic conductivity
$K(h_a)$	cm/d	saturated matrix hydraulic conductivity
K_c	cm/d	hydraulic conductivity accounting for both capillary forces and soil structure
K_f	cm/d	hydraulic conductivity due to adsorption forces
K_s	cm/d	saturated hydraulic conductivity (cm/d)
l	–	a parameter that accounts for tortuosity and connectivity
m	–	m = 1-1/n for VGM model; Fitted parameter for FXW-M3 and B-FXW
n_c	–	free-fitted parameter, when $n_c > 1$ the model is bimodal
n	–	fitted parameter
n_K	–	number of data pairs in the HCC
n_θ	–	number of data pairs in the SWRC
P_k	–	HCC's parameters of hydraulic model
P_θ	–	SWRC's parameters of hydraulic model
P_m	–	the number of model's parameters
$\mathrm{RMSE}_{\log_{10}(K)}$	–	root mean squared error for log-scale hydraulic conductivity
RMSE_θ	cm^3/cm^3	root mean squared error for water content
S_e	cm^3/cm^3	effective water saturation degree related to capillary water
u&v	–	weighting terms that balance the different magnitudes of θ and K
α	1/cm	fitted parameter
θ	cm^3/cm^3	volumetric water content
θ_m	cm^3/cm^3	water content when $h = h_m$
θ_r	cm^3/cm^3	residual soil water content
θ_s	cm^3/cm^3	saturated water content

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Data from Hohenbrink et al. (2023a, 2023b) and Nemes et al. (2015) were used in this manuscript. The code for the developed SPINN-PTFs in this study is publicly available via https://doi.org/10.5281/zenodo.15239681. A user-friendly interface is also provided for users to apply the SPINN-PTFs (https://cug-wang.streamlit.app/).

Acknowledgments
This research was supported in part by the National Natural Science Foundation of China (Number U2244230; 42425207) and in part by the Natural Science Foundation of Hubei (2024AFD376). The authors acknowledge the Editor, Associate Editor, and three anonymous reviewers for their insightful comments that improved the manuscript.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Araya, S. N., & Ghezzehei, T. A. (2019). Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. *Water Resources Research*, *55*(7), 5715–5737. https://doi.org/10.1029/2018WR024357

Báťková, K., Matula, S., Miháliková, M., Hrúzová, E., Abebrese, D. K., Kara, R. S., & Almaz, C. (2023). Prediction of saturated hydraulic conductivity Ks of agricultural soil using pedotransfer functions. *Soil and Water Research*, *18*(1), 25–32. https://doi.org/10.17221/130/2022-SWR

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*, *13*(4), 27–31. https://doi.org/10.1109/45.329294

Beven, K., & Germann, P. (1982). Macropores and water flow in soils. *Water Resources Research*, *18*(5), 1311–1325. https://doi.org/10.1029/WR018i005p01311

Børgesen, C. D., Iversen, B. V., Jacobsen, O. H., & Schaap, M. G. (2008). Pedotransfer functions estimating soil hydraulic properties using different soil parameters. *Hydrological Processes*, *22*(11), 1630–1639. https://doi.org/10.1002/hyp.6731

Børgesen, C. D., Jacobsen, O. H., Hansen, S., & Schaap, M. G. (2006). Soil hydraulic properties near saturation, an improved conductivity model. *Journal of Hydrology*, *324*(1), 40–50. https://doi.org/10.1016/j.jhydrol.2005.09.014

Bottinelli, N., Zhou, H., Boivin, P., Zhang, Z. B., Jouquet, P., Hartmann, C., & Peng, X. (2016). Macropores generated during shrinkage in two paddy soils using X-ray micro-computed tomography. *Geoderma*, *265*, 78–86. https://doi.org/10.1016/j.geoderma.2015.11.011

Botula, Y.-D., Van Ranst, E., & Cornelis, W. M. (2014). Pedotransfer functions to predict water retention for soils of the humid tropics: A review. *Revista Brasileira de Ciência do Solo*, *38*(3), 679–698. https://doi.org/10.1590/S0100-06832014000300001

Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Mining and Knowledge Discovery*, *2*(6), 493–507. https://doi.org/10.1002/widm.1072

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Castellini, M., Di Prima, S., Moret-Fernandez, D., & Lassabatere, L. (2021). Rapid and accurate measurement methods for determining soil hydraulic properties: A review. *Journal of Hydrology and Hydromechanics*, *69*(2), 121–139. https://doi.org/10.2478/johh-2021-0002

Chandrasekhar, P., Kreiselmeier, J., Schwen, A., Weninger, T., Julich, S., Feger, K.-H., & Schwärzel, K. (2018). Why we should include soil structural dynamics of agricultural soils in hydrological models. *Water*, *10*(12), 1862. Article 12. https://doi.org/10.3390/w10121862

Chirico, G. B., Medina, H., & Romano, N. (2010). Functional evaluation of PTF prediction uncertainty: An application at hillslope scale. *Geoderma*, *155*(3), 193–202. https://doi.org/10.1016/j.geoderma.2009.06.008

Coppola, A. (2000). Unimodal and bimodal descriptions of hydraulic properties for aggregated soils. *Soil Science Society of America Journal*, *64*(4), 1252–1262. https://doi.org/10.2136/sssaj2000.6441252x

da Costa, M. B. A., & Cavalcante, A. L. B. (2021). Bimodal soil–water retention curve and *k*-Function model using linear superposition. *International Journal of Geomechanics*, *21*(7), 04021116. https://doi.org/10.1061/(ASCE)GM.1943-5622.0002083

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*(3), 243–253. https://doi.org/10.1016/j.geoderma.2007.11.010

Ding, X., & El-Zein, A. (2024). Predicting soil water retention curves using machine learning: A study of model architecture and input variables. *Engineering Applications of Artificial Intelligence*, *133*, 108122. https://doi.org/10.1016/j.engappai.2024.108122

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. https://doi.org/10.1029/91WR02985

Durner, W. (1994). Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resources Research*, *30*(2), 211–223. https://doi.org/10.1029/93WR02676

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., et al. (2020). Soil structure is an important omission in Earth system models. *Nature Communications*, *11*(1), 522. https://doi.org/10.1038/s41467-020-14411-z

Fredlund, D. G., & Xing, A. (1994). Equations for the soil-water characteristic curve. *Canadian Geotechnical Journal*, *31*(4), 521–532. https://doi.org/10.1139/t94-061

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*(2), 305–319. https://doi.org/10.1029/92WR02339

Gupta, P. K., & Maiti, S. (2023). Enhancing the prediction of hydraulic parameters using machine learning, integrating multiple attributes of GIS and geophysics. *Hydrogeology Journal*, *31*(2), 501–520. https://doi.org/10.1007/s10040-022-02567-5

Gupta, S., Lehmann, P., Bonetti, S., Papritz, A., & Or, D. (2021). Global prediction of soil saturated hydraulic conductivity using random forest in a covariate-based GeoTransfer function (CoGTF) framework. *Journal of Advances in Modeling Earth Systems*, *13*(4), e2020MS002242. https://doi.org/10.1029/2020MS002242

Gupta, S. C., & Larson, W. E. (1979). Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. *Water Resources Research*, *15*(6), 1633–1635. https://doi.org/10.1029/WR015i006p01633

Hadas, A. (1987). Long-term tillage practice effects on soil aggregation modes and strength. *Soil Science Society of America Journal*, *51*(1), 191–197. https://doi.org/10.2136/sssaj1987.03615995005100010040x

Hassan, S. B., Dragonetti, G., Comegna, A., Sengouga, A., Lamaddalena, N., & Coppola, A. (2022). A bimodal extension of the ARYA&PARIS approach for predicting hydraulic properties of structured soils. *Journal of Hydrology*, *610*, 127980.

Hohenbrink, T. L., Jackisch, C., Durner, W., Germer, K., Iden, S. C., Kreiselmeier, J., et al. (2023a). Soil water retention and hydraulic conductivity measured in a widesaturation range. *Earth System Science Data*, *15*(10), 4417–4432. https://doi.org/10.5194/essd-15-4417-2023

Hohenbrink, T. L., Jackisch, C., Durner, W., Germer, K., Iden, S. C., Kreiselmeier, J., et al. (2023b). Soil hydraulic characteristics in a wide range of saturation and soil properties [Dataset]. *GFZ Data Services*. https://doi.org/10.5880/fidgeo.2023.012

Jalabert, S. S. M., Martin, M. P., Renaud, J.-P., Boulonne, L., Jolivet, C., Montanarella, L., & Arrouays, D. (2010). Estimating forest soil bulk density using boosted regression modelling. *Soil Use & Management*, *26*(4), 516–528. https://doi.org/10.1111/j.1475-2743.2010.00305.x

Jarvis, N. J. (2020). 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*, *71*(3), 279–302. https://doi.org/10.1111/ejss.12973

Jena, S., Mohanty, B. P., Panda, R. K., & Ramadas, M. (2021). Toward developing a generalizable pedotransfer function for saturated hydraulic conductivity using transfer learning and predictor selector Algorithm. *Water Resources Research*, *57*(7), e2020WR028862. https://doi.org/10.1029/2020WR028862

Jorda, H., Bechtold, M., Jarvis, N., & Koestel, J. (2015). Using boosted regression trees to explore key factors controlling saturated and near-saturated hydraulic conductivity. *European Journal of Soil Science*, *66*(4), 744–756. https://doi.org/10.1111/ejss.12249

Lebeau, M., & Konrad, J.-M. (2010). A new capillary and thin film flow model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *46*(12), W12554. https://doi.org/10.1029/2010WR009092

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

Lin, H. S., McInnes, K. J., Wilding, L. P., & Hallmark, C. T. (1999). Effects of soil morphology on hydraulic properties: II. Hydraulic pedotransfer functions. *Soil Science Society of America Journal*, *63*(4), 955–961. https://doi.org/10.2136/sssaj1999.634955x

Manyame, C., Morgan, C. L., Heilman, J. L., Fatondji, D., Gerard, B., & Payne, W. A. (2007). Modeling hydraulic properties of sandy soils of niger using pedotransfer functions. *Geoderma*, *141*(3), 407–415. https://doi.org/10.1016/j.geoderma.2007.07.006

McBratney, A., Minasny, B., Cattle, S., & Vervoort, R. (2002). From pedotransfer functions to soil inference systems. *Geoderma*, *109*(1–2), 41–73. https://doi.org/10.1016/S0016-7061(02)00139-8

Minasny, B., Bandai, T., Ghezzehei, T. A., Huang, Y.-C., Ma, Y., McBratney, A. B., et al. (2024). Soil science-informed machine learning. *Geoderma*, *452*, 117094. https://doi.org/10.1016/j.geoderma.2024.117094

Minasny, B., & McBratney, A. B. (2002). The Neuro-m method for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal*, *66*(2), 352–361. https://doi.org/10.2136/sssaj2002.3520

Mohanty, B. P., Bowman, R. S., Hendrickx, J. M. H., & van Genuchten, M. T. (1997). New piecewise-continuous hydraulic functions for modeling preferential flow in an intermittent-flood-irrigated field. *Water Resources Research*, *33*(9), 2049–2063. https://doi.org/10.1029/97WR01701

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *12*(3), 513–522. https://doi.org/10.1029/WR012i003p00513

Muller, K. (2018). Characterising and linking X-ray CT derived macroporosity parameters to infiltration in soils with contrasting structures.

Nemes, A., Schaap, M., Leij, F. J., & Wösten, J. H. M. (2015). UNSODA 2.0: Unsaturated soil Hydraulic database. Database and program for indirect methods of estimating unsaturated hydraulic properties [Dataset]. *US Salinity Laboratory - ARS - USDA*. https://doi.org/10.15482/USDA.ADC/1173246

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*(3–4), 151–162. https://doi.org/10.1016/S0022-1694(01)00465-6

Netto, A. M., Pieritz, R. A., & Gaudet, J. P. (1999). Field study on the local variability of soil water content and solute concentration. *Journal of Hydrology*, *215*(1), 23–37. https://doi.org/10.1016/S0022-1694(98)00259-5

Norouzi, S., Pesch, C., Arthur, E., Norgaard, T., Moldrup, P., Greve, M. H., et al. (2025). Physics-informed neural networks for estimating a continuous form of the soil water retention curve from basic soil properties. *Water Resources Research*, *61*(3), e2024WR038149. https://doi.org/10.1029/2024wr038149

Pachepsky, Y. A., & Rawls, W. J. (2003). Soil structure and pedotransfer functions. *European Journal of Soil Science*, *54*(3), 443–452. https://doi.org/10.1046/j.1365-2389.2003.00485.x

Pachepsky, Y. A., Rawls, W. J., & Lin, H. S. (2006). Hydropedology and pedotransfer functions. *Geoderma*, *131*(3), 308–316. https://doi.org/10.1016/j.geoderma.2005.03.012

Pachepsky, Y. A., Timlin, D., & Varallyay, G. (1996). Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Science Society of America Journal*, *60*(3), 727–733. https://doi.org/10.2136/sssaj1996.03615995006000030007x

Patil, N. G., & Singh, S. K. (2016). Pedotransfer functions for estimating soil hydraulic properties: A review. *Pedosphere*, *26*(4), 417–430. https://doi.org/10.1016/S1002-0160(15)60054-6

Qin, L., Tian, Z., Lin, L., Yi, C., & Chen, J. (2024). Evaluation and development of pedotransfer functions of saturated hydraulic conductivity for subtropical soils. *Geoderma*, *448*, 116976. https://doi.org/10.1016/j.geoderma.2024.116976

Rabot, E., Wiesmeier, M., Schlüter, S., & Vogel, H.-J. (2018). Soil structure as an indicator of soil functions: A review. *Geoderma*, *314*, 122–137. https://doi.org/10.1016/j.geoderma.2017.11.009

Rawls, W., Brakensiek, D., & Logsdon, S. (1993). Predicting saturated hydraulic conductivity utilizing fractal principles. *Soil Science Society of America Journal*, *57*(5), 1193–1197. https://doi.org/10.2136/sssaj1993.03615995005700050005x

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-Fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575. https://doi.org/10.1109/TPAMI.2009.187

Romero-Ruiz, A., Linde, N., Keller, T., & Or, D. (2018). A review of geophysical methods for soil structure characterization. *Reviews of Geophysics*, *56*(4), 672–697. https://doi.org/10.1029/2018RG000611

Rudiyanto, Minasny, B., Chaney, N. W., Maggi, F., Goh Eng Giap, S., Shah, R. M., et al. (2021). Pedotransfer functions for estimating soil hydraulic properties from saturation to dryness. *Geoderma*, *403*, 115194. https://doi.org/10.1016/j.geoderma.2021.115194

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*(3–4), 163–176. https://doi.org/10.1016/S0022-1694(01)00466-8

Schneider, M., & Goss, K.-U. (2012). Prediction of the water sorption isotherm in air dry soils. *Geoderma*, *170*, 64–69. https://doi.org/10.1016/j.geoderma.2011.10.008

Selahvarzi, M., Naghedifar, S. M., Oliazadeh, A., & Loáiciga, H. A. (2025). Hierarchical pseudo-continuous machine-learning-based pedotransfer models for infiltration curves: An investigation on the role of regularization and ensemble modeling. *Journal of Hydrology*, *650*, 132459. https://doi.org/10.1016/j.jhydrol.2024.132459

Shekhar, S., Amin, M. B., & Khandelwal, P. (1992). Generalization performance of feed-forward neural networks. In E. Gelenbe (Ed.), *Neural networks* (pp. 13–38). North-Holland. https://doi.org/10.1016/B978-0-444-89330-7.50005-3

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., & Pásztor, L. (2019). Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrology and Earth System Sciences*, *23*(6), 2615–2635. https://doi.org/10.5194/hess-23-2615-2019

Szabó, B., Weynants, M., & Weber, T. K. D. (2021). Updated European hydraulic pedotransfer functions with communicated uncertainties in the predicted variables (euptfv2). *Geoscientific Model Development*, *14*(1), 151–175. https://doi.org/10.5194/gmd-14-151-2021

Toth, B., Weynants, M., Nemes, A., Mako, A., Bilas, G., & Toth, G. (2015). New generation of hydraulic pedotransfer functions for Europe. *European Journal of Soil Science*, *66*(1), 226–238. https://doi.org/10.1111/ejss.12192

Tuller, M., & Or, D. (2001). Hydraulic conductivity of variably saturated porous media: Film and corner flow in angular pore space. *Water Resources Research*, *37*(5), 1257–1276. https://doi.org/10.1029/2000wr900328

Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, *44*(5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., et al. (2017). Pedotransfer functions in Earth system science: Challenges and perspectives. *Reviews of Geophysics*, *55*(4), 1199–1256. https://doi.org/10.1002/2017RG000581

Vereecken, H., Amelung, W., Bauke, S. L., Bogena, H., Brüggemann, N., Montzka, C., et al. (2022). Soil hydrology in the Earth system. *Nature Reviews Earth and Environment*, *3*(9), 573–587. https://doi.org/10.1038/s43017-022-00324-6

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*(6), 389–403. https://doi.org/10.1097/00010694-198912000-00001

Wang, Y., Jin, M., & Deng, Z. (2018). Alternative model for predicting soil hydraulic conductivity over the complete moisture range. *Water Resources Research*, *54*(9), 6860–6876. https://doi.org/10.1029/2018WR023037

Wang, Y., Ma, J., & Guan, H. (2016). A mathematically continuous model for describing the hydraulic properties of unsaturated porous media over the entire range of matric suctions. *Journal of Hydrology*, *541*, 873–888. https://doi.org/10.1016/j.jhydrol.2016.07.046

Wang, Y., Ma, R., & Vereecken, H. (2025). A generalized framework to describe unimodal and bimodal soil hydraulic properties over full water saturation range. *Water Resources Research*, *61*(2), e2024WR038450. https://doi.org/10.1029/2024WR038450

Wang, Y., Ma, R., & Zhu, G. (2022). Improved prediction of hydraulic conductivity with a soil water retention curve that accounts for both capillary and adsorption forces. *Water Resources Research*, *58*(4), e2021WR031297. https://doi.org/10.1029/2021WR031297

Wang, Y., Ma, R., & Zhu, G. (2023). Representation of the influence of soil structure on hydraulic conductivity prediction. *Journal of Hydrology*, *619*, 129330. https://doi.org/10.1016/j.jhydrol.2023.129330

Wang, Y., Zhou, J., Ma, R., Zhu, G., & Zhang, Y. (2022). Development of a new pedotransfer function addressing limitations in soil hydraulic models and observations. *Water Resources Research*, *58*(6), e2021WR031406. https://doi.org/10.1029/2021WR031406

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*(9), e2019WR026820. https://doi.org/10.1029/2019WR026820

Weber, T. K. D., Weihermueller, L., Nemes, A., Bechtold, M., Degre, A., Diamantopoulos, E., et al. (2024). Hydro-pedotransfer functions: A roadmap for future development. *Hydrology and Earth System Sciences*, *28*(14), 3391–3433. https://doi.org/10.5194/hess-28-3391-2024

Weynants, M., Vereecken, H., & Javaux, M. (2009). Revisiting vereecken pedotransfer functions: Introducing a closed-form hydraulic model. *Vadose Zone Journal*, *8*(1), 86–95. https://doi.org/10.2136/vzj2008.0062

Wösten, J., Lilly, A., Nemes, A., & Le Bas, C. (1998). Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning (Vol. *156*). Report.

Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimizationb. *Journal of Electronic Science and Technology*, *17*(1), 26–40. https://doi.org/10.11989/JEST.1674-862X.80904120

Youngs, E. G. (1988). Soil physics and hydrology. *Journal of Hydrology*, *100*(1), 411–431. https://doi.org/10.1016/0022-1694(88)90194-1

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

Zhang, Y., & Schaap, M. G. (2019). Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. *Journal of Hydrology*, *575*, 1011–1030. https://doi.org/10.1016/j.jhydrol.2019.05.058

Zhang, Y., Schaap, M. G., & Zha, Y. (2018). A high-resolution global map of soil hydraulic properties produced by a hierarchical parameterization of a physically based water retention model. *Water Resources Research*, *54*(12), 9774–9790. https://doi.org/10.1029/2018WR023539

Zhang, Y., Wang, L., Zhang, W., Zhang, Z., & Zhang, M. (2023). Quantification of root systems and soil macropore networks association to soil saturated hydraulic conductivity in forested wetland soils. *Forests*, *14*(1), 132. Article 1. https://doi.org/10.3390/f14010132

Zhang, Y., Weihermüller, L., Toth, B., Noman, M., & Vereecken, H. (2022). Analyzing dual porosity in soil hydraulic properties using soil databases for pedotransfer function development. *Vadose Zone Journal*, *21*(5), e20227. https://doi.org/10.1002/vzj2.20227