Limitations of the Model Structure and Soil Hydraulic Property Observations in Pedotransfer Function Development

Yunquan Wang[1,*], Jieliang Zhou[1], Rui Ma[1], Gaofeng Zhu[2], Yongyong Zhang[3]

[1] *Hubei Key Laboratory of Yangtze River Basin Environmental Aquatic Science*, School of Environmental Studies, China University of Geosciences at Wuhan, 430074, PR China,

[2]Key Laboratory of Western China's Environmental Systems (Ministry of Education), Lanzhou University, Lanzhou 730000, China,

[3]Linze Inland River Basin Research Station, Key Laboratory of Ecohydrology of Inland River Basin,

Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

*Corresponding Author: Yunquan Wang, School of Environmental Studies, China University of Geosciences at Wuhan, Lumo Rd. 388, Hongshan District, Wuhan, China, 430074 (wangyq@cug.edu.cn)

Abstract

The commonly applied pedotransfer functions (PTFs), which predict soil hydraulic properties (SHPs) from easily measured soil properties such as texture information, often account only for capillary forces. Recent advances in soil hydraulic modeling suggest that, to improve the prediction of SHPs under dry conditions, the impact of adsorption forces has to be taken into account. However, the lack of observations in particularly dry conditions, due to the difficult and time-consuming measurement, hinders the development of PTFs that predict SHPs from saturation to oven dryness. In this paper, we first present a simple method for predicting complete SHPs with limited measurements that cover only a relatively high matric potential range. With this method, we extended a public dataset to cover dry conditions, and then applied it to develop PTFs that can predict SHPs from saturation to oven dryness. This was achieved by applying the complete soil hydraulic model proposed by Wang et al. (2021), which accounts for both capillary and adsorptions forces and overcomes the unrealistic decrease near saturation for fine-textured soils. The impact of vapor diffusion was also considered. We further applied this method in extending an existing capillary-based PTF to dry conditions. The results showed that: 1) the proposed method performs very well in describing SHPs over the entire moisture range; 2) the PTFs developed with the extended observations and the complete model show a superior prediction performance, especially for the hydraulic conductivity; and 3) the extended capillary-based PTF improves the performance in describing SHPs under dry conditions.

1. Introduction

Soil hydraulic properties (SHPs), including the soil water retention curve (SWRC) and the hydraulic conductivity curve (HCC), are crucial parameters

in water and solute transport simulation. However, the experimental determination of these parameters is expensive, time-consuming, and can be difficult, especially under very dry conditions (Vereecken et al., 2010). For large-scale applications, the experimental measurements can even be impractical, given the requirement for a large number of measurements. As a result, pedotransfer functions (PTFs), which relate SHPs to easily measured soil properties such as soil texture, serve as a standard method for predicting SHPs, especially in large-scale field-based applications (e.g., Zhang et al., 2018; Dai et al., 2019).

The general way of developing a PTF is fitting the measured SHPs in a dataset with soil hydraulic models to obtain the parameters, and then developing the relationships between the parameters and the easily measured soil properties such as soil texture through the use of regression methods. Over the past few decades, many PTFs have been developed, and great efforts have been made to improve their performance, mainly through including large measured datasets (e.g., Wösten et al., 1999; Nemes et al., 2001; Weynants et al., 2013; Tóth et al., 2015), adding additional soil properties such as organic content or soil chemistry information as the input (e.g., Rawls & Pachepsky, 2002; Pachepsky & Rawls, 2004; Børgesen & Schaap, 2005; Pachepsky et al., 2006; Børgesen et al., 2008; Tóth et al., 2015; Szabó et al., 2021), and applying more powerful machine learning methods such as nearest neighbor methods, support vector machine, or random forest (RF) (e.g., Nemes et al., 2006; Lamorski et al., 2008; Araya & Ghezzehei, 2019; Szabó et al., 2021). Useful reviews of PTF development can be found in Wösten et al. (2001), Pachepsky and Rawls (2004), Vereecken et al. (2010), and Van Looy et al. (2017), among others.

However, almost all the soil hydraulic models applied for developing PTFs are capillary-based. For example, the most widely applied model is the well-known van Genuchten (1980)-Mualem model (1976) (hereafter referred to as the VGM model). While these models have a good ability to describe SHPs in the high to medium moisture range, the failure of these capillary-based models has long been recognized under low water content conditions (e.g., Nimmo, 1991; Rossi & Nimmo, 1994; Tuller & Or, 2001; Wang et al., 2013). The reason for this is that these models do not consider the impact of adsorption forces, which are dominant in a low water content situation (e.g., Tuller & Or, 2001; Tokunaga, 2009; Wang et al., 2016; 2018; Chen et al., 2017). As drylands cover nearly 41.3% of the land surface (Robinson, 2015), understanding soil water dynamics, especially in the dry moisture range, as well as their impact on evapotranspiration and other related processes, is crucial for the understanding of the global water and energy cycles and their response to climate change. For example, in a recent study by Wang et al. (2019), they argued that adsorption forces have a significant impact on soil evaporation estimation. The simulation and prediction of these related processes requires the accurate determination of SHPs that cover dry conditions, where the adsorption forces play a crucial role.

Early efforts were mainly focused on developing a complete SWRC over the entire moisture range (e.g., Nimmo, 1991; Rossi & Nimmo, 1994; Fredlund &

Xing, 1994; Fayer & Simmons, 1995; Webb, 2000; Lu et al., 2008). Tuller and Or (2001) were the first to include non-capillary forces when developing a complete HCC. Since then, a series of complete models have been proposed that consider the coupled effects of capillary and adsorption forces, which have performed well in describing SHPs from saturation to oven dryness (e.g., Lebeau & Konrad, 2010; Zhang, 2011; Wang et al., 2016; Liao et al., 2018; Stanić et al., 2020 among others). However, to describe a complete HCC, an additional parameter that represents the saturated film conductivity is often required, which is difficult to determine. Differing from these developed models that treat the total conductivity as a sum of the capillary conductivity and the non-capillary conductivity, Wang et al. (2018) proposed a new model to describe the HCC over the entire moisture range, using a single equation. Compared to the commonly used capillary-based models (such as the VGM model), this model requires no extra parameters, and shows a very good ability in predicting the HCC. Together with the SWRC described by the Fredlund and Xing (1994) model, this represents an easy way to develop new PTFs that can predict SHPs over the entire moisture range. For example, by applying this Fredlund and Xing (1994)-Wang et al. (2018) model (referred to as the FXW model hereafter), Rudiyanto et al. (2021) recently developed a new PTF and achieved a significant reduction in root-mean-square error for both the SWRC and HCC, in comparison with the VGM model based Rosetta3 PTF developed by Zhang and Schaap (2017) and another complete PTF proposed by Weber et al. (2020). The original FXW model, however, does have one limitation, i.e., the HCC drops dramatically near saturation for soils with small n value, which is a parameter used in shaping the SWRC (Wang et al., 2018; de Rooij et al., 2021). Wang et al. (2021) recently improved the FXW model further to overcome this shortcoming by introducing a non-zero air-entry value.

Besides the limitations of the model structure, another crucial limitation in developing PTFs that aim to predict SHPs from saturation to oven dryness comes from the limited measurements. That is, the applied soil hydraulic data for PTF development often do not include measurements from very dry conditions. For example, most of the SWRC measurements in the UNsaturated SOil hydraulic DAtabase (UNSODA) adopted by Rudiyanto et al. (2021) are for a matric potential of higher than about -1.0×10^3 cm. Lu et al. (2014) showed that the observed data should cover a more negative potential range to achieve an accurate description of the SWRC over the entire moisture range. When it comes to the HCC, the measurements under dry conditions are even fewer. Specifically, only 29 samples in the UNSODA database cover the conductivity measurements at a matric potential of higher than -1.0×10^4 cm. Since the number and the quality of the measured data are the key points for developing PTFs, the lack of measurements under low moisture conditions significantly limits the reasonable development of PTFs that predict SHPs over the complete moisture range. This limitation comes from the SHP observations, which, to the best of our knowledge, is an aspect that has rarely been considered or dealt with in the literature.

Meanwhile, many of the existing PTFs that only consider capillary forces were developed on different, usually non-public datasets (e.g., Tóth et al., 2015; Szabó et al., 2021). Thus, it would be beneficial to develop a simple way to predict SHPs from saturation to oven dryness with these existing PTFs.

Accordingly, the purpose of this study was: 1) to develop a method for predicting SHPs over the entire moisture range with limited observations that cover only a relatively high potential range; 2) to develop new PTFs that account for both capillary and non-capillary effects with the application of this extended method; and 3) to provide an easy way to extend the existing PTFs to dry conditions.

2. Methods and Materials

2.1. The soil hydraulic model that accounts for both capillary and non-capillary effects

In this study, the FXW model and its modification proposed recently by Wang et al. (2021) were applied to describe the SHPs over the entire moisture range.

2.1.1. The original FXW model

The original FXW model accounts for the impact of both capillary and adsorption forces, and requires no additional parameters, compared to the commonly used capillary models, such as the well-known VGM model.

The SWRC of the FXW model comes from Fredlund and Xing (1994), and is written as:

where $S = \theta / \theta_s$ is the saturation degree, with θ ($L^3\ L^{-3}$) being the volumetric water content and θ_s ($L^3\ L^{-3}$) the saturated water content; h (L) is the matric potential; h_r, with a typical value of -1.5×10^3 cm, is simply explained as a shaping factor; and h_0, which was set as -6.3×10^6 cm, according to the suggestion of Schneider and Goss (2012), is the matric potential corresponding to a water content value of 0. $\Gamma(h)$ is expressed as:

where α (L^{-1}), n, and m are all the fitted parameters, and e is the Euler's number (the e constant).

The HCC of the FXW model that accounts for liquid flow was developed by Wang et al. (2018), and is expressed as:

where K_l ($L\ T^{-1}$) and K_s ($L\ T^{-1}$) are the liquid conductivity and the saturated hydraulic conductivity, respectively; and l is the fitting parameter, which has a typical value of 3.5, as suggested by Wang et al. (2018).

2.1.2. The modified FXW-M model

The HCC of the original FXW model, however, drops dramatically near saturation for n values close to 1 (Wang et al., 2018; de Rooij et al., 2021). This comes from the non-zero $d\theta/dh$ at the matric potential of zero (van Genuchten & Nielsen, 1985; Schaap & van Genuchten, 2006; de Rooij et al., 2021). To overcome this limitation, Wang et al. (2021) recently proposed a modified HCC

by introducing a non-zero air-entry value h_s, following Vogel et al. (2000) and Ippisch et al. (2006). The new HCC is written as:

where h_s is -0.2 cm, as suggested by Wang et al. (2021).

The new HCC model described with equation (4), together with the SWRC described in equation (1), are termed the FXW-M model. For a detailed description of the FXW-M model, we refer the reader to Wang et al. (2021).

2.1.4. The impact of vapor diffusion

In this study, since vapor diffusion also contributes to the total water flux under very dry conditions, the isothermal vapor diffusion was also considered. Following Saito et al. (2006), the isothermal conductivity of vapor flow can be written as:

where $_v$ (kg m^{-3}) is the saturated vapor density, and D_v (m^2 s^{-1}) is the vapor diffusivity in soil, which can be written as:

where $_a$ is the air-filled porosity; is the tortuosity factor calculated according to Millington and Quirk (1961); and D_a (m^2 s^{-1}) is the vapor diffusivity in air, which can be expressed as:

where T (K) is the absolute temperature.

The total conductivity K (L T^{-1}) is therefore written as:

2.2. Predicting SHPs over the entire moisture range with limited observations

2.2.1. The SWRC

Measured soil hydraulic data are rare at a very low matric potential, limiting the application of a complete model. Fortunately, it is well-known that the SWRC, which is controlled mainly by the adsorption forces under dry conditions, can be described by a semi-log scale linear relationship (Campbell & Shiozawa, 1992). This is generally in the form of:

with *SL* being the dimensionless slope.

The soil water content and then *SL* are determined by the specific soil surface area, which is, in turn, controlled mainly by the clay fraction. Several relationships between the clay fraction and *SL* can be found in the literature (e.g., Resurreccion et al., 2011; Schneider & Goss, 2012; Arthur et al., 2013). Jensen et al. (2015) also included the impact of the silt fraction and organic matter in deriving *SL*.

However, the referred relationships were usually built through simple regression with the application of limited datasets. For example, Resurreccion et al. (2011) developed an exponential relationship between *1/SL* and the clay fraction based on 41 soil samples, and Schneider and Goss (2012) provided a linear relationship between these two parameters for 18 soil samples.

In this study, by collecting a total of 275 soil samples from the literature (as described in Section 2.7), we built a new relationship between *SL* and the soil texture information, through the RF model, which is a powerful machine learning method.

With this developed method, the SWRC under dry conditions can thus be predicted from the soil texture information, including the sand, silt, and clay percentages, and the bulk density. It should be noted that, for the applied datasets, the measurements were generally performed in very dry conditions, where the van der Waals forces dominate (e.g., Resurreccion et al., 2011; Schneider & Goss, 2012). Therefore, an upper boundary of about -1.0×10^5 cm for the matric potential was set when applying this method (Tuller & Or, 2005).

In this study, three additional water retention data points at the matric potential of -1.0×10^5 cm, -5.0×10^5 cm, and -1.0×10^6 cm, respectively, were predicted with the derived method. Together with the observations at a relatively high matric potential range, they were fitted with the SWRC, as described in equation (1), to obtain the parameters.

2.2.2. The HCC

The hydraulic conductivity, which accounts for the adsorption forces, is determined by the specific surface area S_A (L^2 L^{-3}) and the film thickness f (Bird, 1960). A detailed description of a method for predicting conductivity data from the SWRC was presented in a recent study by Wang et al. (2021). This is briefly introduced as follows.

The conductivity accounting for the adsorption forces can be expressed as (Wang et al., 2017; Lebeau & Konrad, 2010):

where is the water density (9.98×10^2 kg m^{-3}), g is the acceleration of gravity (9.81 m s^{-2}), and is the fluid viscosity (1.005×10^{-3} Pa s at 293 K). $B(f)$ is the correction factor that accounts for the modified viscosity for a film thickness of thinner than 10 nm (Or & Tuller, 2000; Lebeau & Konrad, 2010), which is expressed as:

where a is 5.53×10^{-10} m at 293 K, and $\mathrm{Ei}(-x) = -\int_x^\infty \left[\frac{\exp(-t)}{t}\right] \mathrm{dt}$ is the exponential integral.

The film thickness f is controlled by both the electrostatic forces (Langmuir, 1938; Tokunaga, 2009; 2011) and the van der Waals forces (Iwamatsu & Horii, 1996), and is expressed as:

where is the relative permittivity of water (78.54); $_0$ is the permittivity of free space (8.85×10^{-12} C^2 J^{-1} m^{-1}); k_B is the Boltzmann constant (1.381×10^{-23} J K^{-1}); T is the temperature in kelvins; z is the ion valence, which was set to 1 in this study; and e_c is the electron charge (1.602×10^{-19} C). A_{svl} is the Hamaker constant for solid-vapor interactions, which was set to -6.0×10^{-20} J, following Tuller and Or (2005).

The S_A in equation (9) can be approximately estimated by dividing the soil water content by the film thickness, as suggested by Tuller and Or (2005). By taking a typical matric potential h_m, the specific surface area can be estimated as:

By substituting all the parameters, the calculated hydraulic conductivity by equation (10) would be:

where $b(h_m)$ is 1.084×10^{-4} cm d^{-1}, 8.678×10^{-6} cm d^{-1}, and 2.693×10^{-6} cm d^{-1} at the three selected potentials of -1.0×10^4 cm, -5.0×10^4 cm, and -1.0×10^5 cm, respectively.

2.3. Parameter optimization

With the extended SHP observations, the soil hydraulic model is optimized to derive the parameters. The objective function to be minimized is defined as:

where n and n_K are the number of data pairs for the retention and the conductivity function, respectively; $_i$ and K_i are the measured water content and hydraulic conductivity, respectively; and and are the model-estimated values, respectively. We set $w = 1000$ and $w_k = 4$ as the weights of the water content and conductivity data. Note that, for datasets with only the SWRC, w_k should be set to 0. $b = [log_{10}\ (\ ),\ log_{10}(n),\ log_{10}\ (m),\ _s,\ log_{10}\ (K_s)]$ is the parameter vector used for the optimization, and l was set to 3.5, as suggested by Wang et al. (2018).

For the parameter optimization, an upper boundary of 1.5 was set for m for all of the models, as suggested by Wang et al. (2016). For parameter n, a low boundary of 1.1 was set for the original FXW model (Rudiyanto et al., 2021), while a value of 1.01 was set for the FXW-M model. The higher low boundary limit of n for the FXW model was set to avoid the dramatic decrease of the HCC (Wang et al., 2018). The optimization was achieved by applying the shuffled complex evolution (SCE-UA) method developed at the University of Arizona, as proposed by Duan et al. (1992).

2.4. Developing PTFs through the random forest model

With the known soil texture information and the optimized parameters, PTFs that can predict SHPs from saturation to oven dryness can be developed.

In this study, we developed three PTFs in regard to the input information. Model H1 represents the model input of the soil class, model H2 accounts for the input of the sand, silt, and clay percentages, while model H3 has an additional input of bulk density. For models H2 and H3, both the original data and the extended data obtained with the method described in Section 2.2 were applied. Because the observations were generally sufficient for describing the SHPs over the entire moisture range, when the measurements reached the potential of about -1.0 $\times$ 10^4 cm (Lu et al., 2014), only the datasets with measurements above this potential were extended.

For developing PTFs, the RF model was adopted. The RF model (Breiman, 2001) is regarded as one of the best machine learning techniques (e.g., Cutler et al., 2007; Boulesteix et al., 2012). Predictions in RF are generated as an ensemble estimate by constructing a lot of decision trees through bootstrap samples (Hengl et al., 2018). RF is easy to adopt by the use of a package such as the Ranger package (Wright & Wager, 2017) implemented in R software. The RF model has been successfully applied in predicting soil properties (e.g., Hengl et al., 2017; Araya & Ghezzehei, 2019; Szabó et al., 2021).

When applying the RF model, the number of trees was set to 500, which enabled a stable root-mean value for the predictions. The minimum leaf size was derived by optimization, and a low boundary of 5 was set for all the parameters.

2.5. Extending the existing capillary-based PTFs to predict SHPs over the entire moisture range

With the extended method, as described in Section 2.2, the existing PTFs that account for capillary forces can be easily extended to predict SHPs over the entire moisture range.

In this study, we took the Rosetta3 PTF proposed by Zhang and Schaap (2017), which was developed based on the VGM model that accounts for only capillary forces, as an example. Firstly, the SHPs that were dominated by capillary forces were predicted with Rosetta3. This low boundary for the capillary-dominant zone is defined by a critical matric potential h_c that marks the end of capillary continuity, which can be expressed following Lehmann et al. (2008) as:

where $_{vG}$ and m_{vG} are the parameters of the VGM model (Appendix A).

Secondly, the predicted SHPs from 0 to h_c, together with the extended SHPs, as calculated by equations (9) and (14), were fitted with the FXW model to derive the parameters, which can be applied to describe the SHPs from saturation to oven dryness.

2.6. Model performance statistics

The root-mean-square error ($RMSE$), mean error (ME), and coefficient of determination (R^2) were applied for evaluating the model performance.

The $RMSE$ is defined as:

where N represents the number of data pairs, and o_i and $\hat{o}_i$ are the measured and estimated values, respectively.

The ME for quantifying systematic errors is defined as:

The R^2 is defined as:

where $\overline{o}$ is the mean value of o_i.

2.7. Data

The UNSODA database (Nemes et al., 2001) was applied to develop the PTFs. Only the measurements from drying experiments in the laboratory were chosen, to avoid the impact of hysteresis. For the data selection, a low boundary of 1.0 g cm cm^{-3} was set for the bulk density, and an upper boundary of 0.65 cm^3 cm^{-3} was set for the saturated water content. In addition, a low boundary of -300 cm for the matric potential, which is a value close to the so-called field capacity, was set for the SWRC data to ensure the representation of the measurements. Considering all the limitations, a total of 422 soil samples were selected, with a total of 4,887 retention points. Among the samples, 215 soil samples also included HCC measurements, with a total of 3,966 points. The soil texture distribution of the selected soil samples is shown in Figure 1. In order to test the performance of the PTFs, the selected data were randomly divided into two groups, i.e., a training set with 70% of the data and a test set with the remaining 30% of the data.

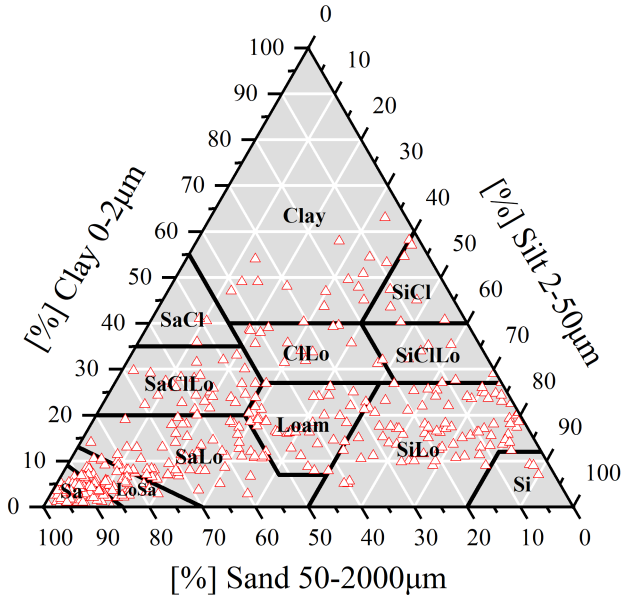

Figure 1. Soil texture distribution of the 422 selected soil samples from the UNSODA database

To build the relationship between SL in equation (9) and the soil texture information, 62 samples from Resurreccion et al. (2011) and Jensen et al. (2015) were selected. In addition, 213 soil samples from the UNSODA database with a minimum water potential of less than -1.0×10^4 cm were selected to derive

the *SL* through inverse modeling. This was achieved by optimizing the SWRC with the FXW model to obtain the parameters, and then predicting the water content values at a matric potential of -1.0×10^5 cm, -5.0×10^5 cm, and -1.0×10^6 cm. The *SL* was then derived by fitting equation (9) with these predicted water retention points.

3. Results

3.1. Prediction of the complete SHPs with limited observations

3.1.1. Determination of *SL* from the soil texture information

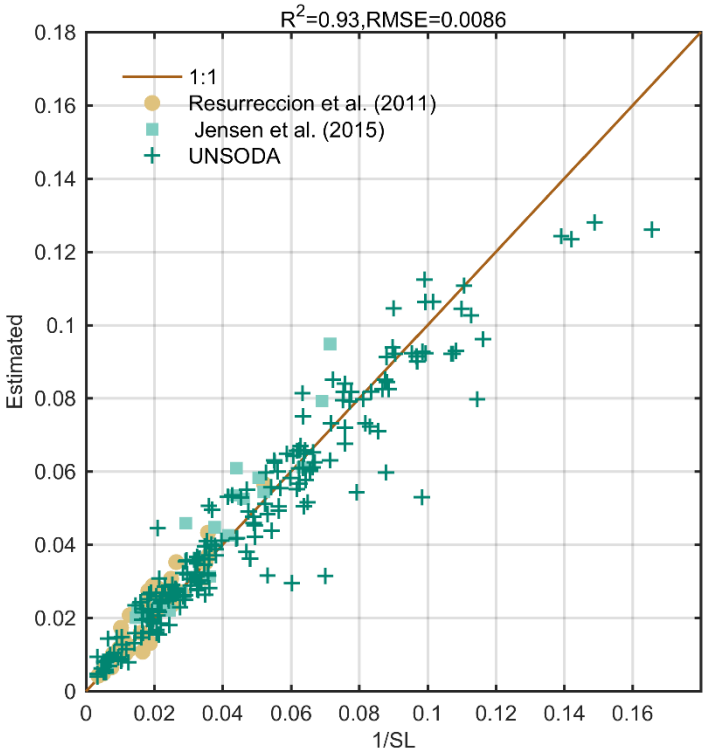

Figure 2. Prediction of *SL* with the random forest method

Figure 2 presents the predicted *SL* obtained with the RF method. A total of 275 soil samples, with a variety of soil textures, were used for the testing. As shown, the *SL* in equation (9) can be well predicted with the soil texture information by applying the RF method, particularly for *1/SL* values of less than 0.04. A higher *SL* generally means a higher clay fraction in the soil sample. For more fine-textured soils (with much higher *1/SL* values), the method underestimates the *1/SL* value for some of the UNSODA database. It should be noted that, for the UNSODA database, the *1/SL* values were derived from inverse modeling rather than direct measurement. As a result, there is higher uncertainty than in the data from Resurreccion et al. (2011) and Jensen et al. (2015). The overall

performance of the *1/SL* estimation is very good, with an *RMSE* of 0.0086 and an R^2 of 0.93. The required inputs are the sand, silt, and clay percentages, as well as the bulk density.

3.1.2. Prediction of a complete SWRC with limited observations

With the *SL* predicted from the soil texture information, three water retention points at a matric potential of -1.0×10^5 cm, -5.0×10^5 cm, and -1.0×10^6 cm could be estimated with equation (9). These three points, together with the measurements at a relatively high matric potential range, were then fitted with the FXW model. As shown in Figure 3, this extended method generally yields a good performance in describing the SWRC from saturation to oven dryness. The observations at the non-fitted range are captured well by the proposed model, especially when the measurements reach the potential range of about -1.0×10^3 cm. The *RMSE* is 0.017 cm^3 cm^{-3}, which is very close to the 0.013 cm^3 cm^{-3} of the fitted results with all the measurements (Figure 3a). The model performances with measurements higher than -100 cm and -300 cm are also presented in Figure 3. These two situations also show good agreement with the observations for a water content of less than about 0.2 cm^3 cm^{-3}, whereas there is underestimation for some soil samples in the wet moisture range where capillary forces dominate.

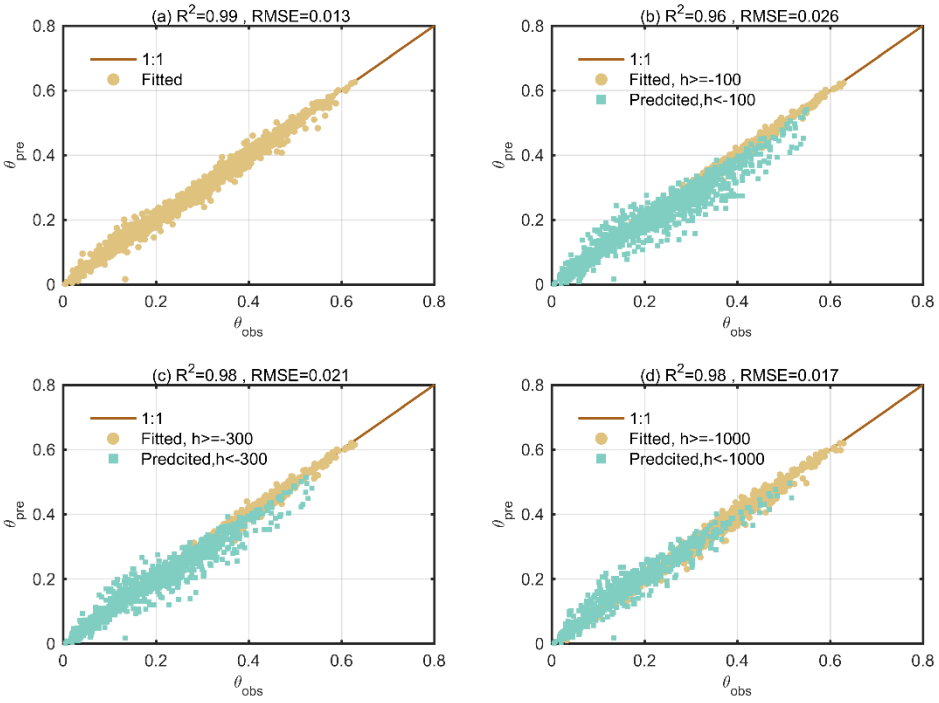

Figure 3. Prediction of the complete SWRC with limited observations. The fitting results in (a) represent the results obtained with all the observations, while those in (b), (c), and (d) represent the fitted results obtained with only measurements for a matric potential higher than −100 cm, −300 cm, and −1000 cm, respectively, as well as the three additional water retention points predicted with equation (9) in dry conditions.

In contrast, the fitted results obtained without the extended retention points show significant overestimation of the water content in dry conditions, with roughly double the *RMSE* compared to those results obtained with the extended method (Figure S1 in the supplementary material).

3.1.3. Prediction of a complete HCC with limited observations

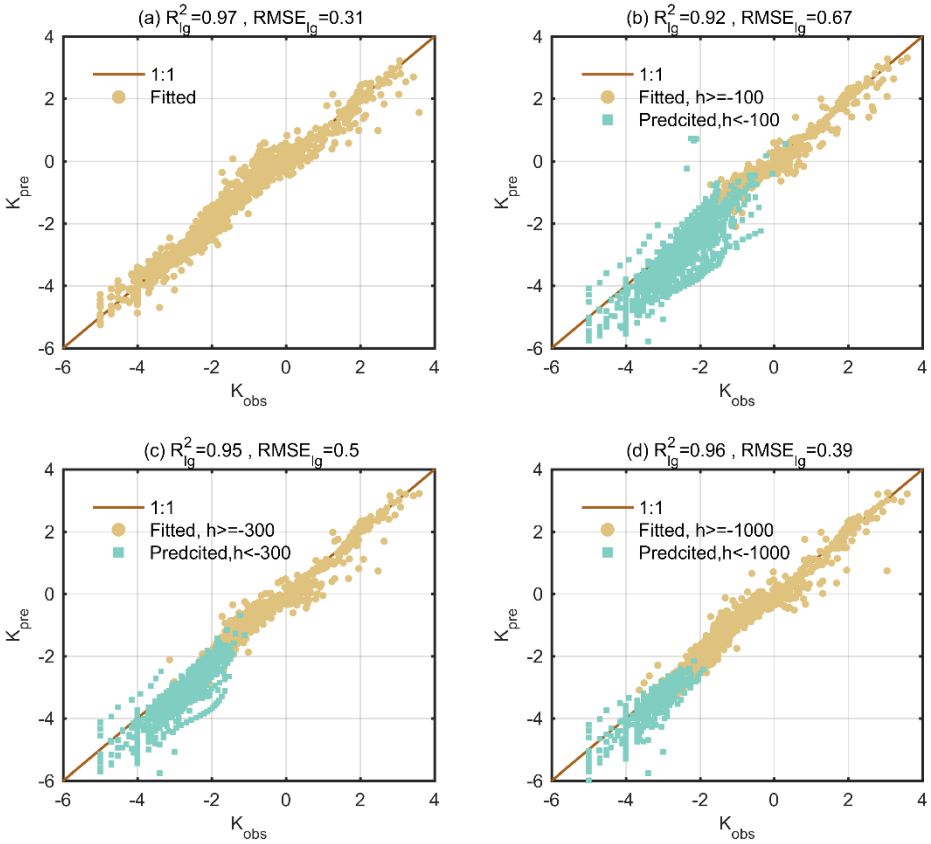

Figure 4. Prediction of a complete HCC with limited observations. The fitting results in (a) represent the results obtained with all the observations, while those in (b), (c), and (d) represents the fitted results obtained with only measurements for a matric potential higher than −100 cm, −300 cm, and −1000 cm, respec-

tively, as well as the three additional hydraulic conductivity points predicted with equation (14) in dry conditions.

Figure 4 indicates that the extended method, as presented in Section 2.2.2, performs well in predicting the hydraulic conductivity under dry conditions. For observations with a low potential boundary of −100 cm and −300 cm, the $RMSE_{log10(K)}$ is 0.67 and 0.50 cm d^{-1}, respectively. When the observations reach a potential range of -1.0×10^3 cm, the extended method yields a much better performance (Figure 4d), with the $RMSE_{log10(K)}$ being 0.39 cm d^{-1}, which is much closer to the 0.31 cm d^{-1} of the fitted curve obtained with all the observations. In contrast, the predictions obtained without the extended data yield a much poorer performance under dry conditions, especially for the curve with observations higher than −100 cm, with a lower R^2 of 0.85 and a much higher $RMSE_{log10(K)}$ of 0.80 cm d^{-1} (Figure S2 in the supplementary material).

3.2. Development of PTFs with the extended and original data

3.2.1. Model H1

Following the standard of the United States Department of Agriculture, 12 soil classes were categorized based on the percentages of the sand, silt, and clay content. For each soil class, we calculated the mean value and the standard deviation of the five parameters of the FXW-M model (Table 1). Here, we only provide the mean values derived with the extended data.

It should be noted that, because the particle size of some samples was located at the boundary of two different classes, these samples accounted for both soil classes, resulting in the total number of samples exceeding 422. As shown, some soil classes only cover a few soil samples, so caution should be applied when applying the mean values of these soil classes.

Table 1. The derived mean value and the standard deviation (SD) of the five parameters of the FXW-M model for the 12 soil classes.

Soil texture	Num[a]	α (1/cm)		n		m	s (cm^3cm^{-3})		$log_{10}K_s$[b]
		Mean	SD	Mean	SD	Mean	SD		Mean
Clay	16(12)	0.10	0.09	2.05	1.49	0.18	0.09		0.53
Clay loam	15(4)	0.08	0.10	1.68	1.01	0.26	0.13		0.45
Loam	50(33)	0.08	0.09	1.73	1.63	0.37	0.15		0.47
Loamy sand	48(16)	0.04	0.02	3.32	2.12	0.72	0.32		0.39
Sand	99(48)	0.04	0.03	4.62	2.79	0.84	0.30		0.36
Sandy clay	3(0)	0.00	0.00	2.20	1.99	0.26	0.16		0.41
Sandy clay loam	29(5)	0.05	0.07	2.08	2.21	0.34	0.16		0.39
Sandy loam	63(24)	0.03	0.04	2.20	1.97	0.58	0.22		0.37
Silt	3(3)	0.01	0.00	1.86	1.19	0.74	0.27		0.42
Silty loam	86(58)	0.04	0.07	1.41	1.02	0.56	0.26		0.44
Silty clay	8(8)	0.08	0.07	2.50	1.37	0.12	0.06		0.48

Soil texture	Num[a]	(1/cm)	n	m	s (cm^3cm^{-3})	$log_{10}K_s$[b]
Silty clay loam	9(4)	0.08	0.09	2.74	3.02	0.29	0.16	0.49

a. the number in brackets is for conductivity. **b**. Nan means no dataset

3.2.2. Models H2 and H3

For the original FXW model and its modified form, the training and test performance of the parameters, represented by the R^2, *RMSE*, and *ME*, are presented in Tables 2–3, for both models H2 and H3.

In the training case, the FXW model, together with the RF method, performs very well. Taking model H3 as an example, the PTFs developed with the original data have R^2 values of higher than 0.75, the *RMSE* varies from 0.04 for $_s$ to 0.60 for $log_{10}(K_s)$, and the *ME* has a magnitude of about 1.0×10^{-3} for all five parameters. The test case shows a comparable performance with the training case in predicting the four parameters for describing the SWRC, with slightly lower R^2 and higher *RMSE* values (Tables 2 and 3). The worst performance is for parameter $log_{10}(\ )$, with the highest *RMSE* of 0.45 cm^{-1}. When it comes to parameter $log_{10}(K_s)$, the test case of the FXW model, however, yields a much poorer performance. The R^2 decreases from 0.79 to 0.22, the *RMSE* increases from 0.60 to 0.87 cm d^{-1}, and the *ME* increases from -0.003 to 0.015 cm d^{-1} when compared to the training case (model H3). In contrast, the FXW-M model significantly improves the prediction of $log_{10}(K_s)$ in the test case, with the statistical values close to those of the training case.

Compared to the PTFs developed with the original data, the PTFs developed with the extended data show a similar performance in predicting the four parameters of the SWRC for both the training and test cases. The PTFs show a slight improvement in predicting parameters $log_{10}(\ )$ and $log_{10}(m)$, and a reduced performance for parameter $log_{10}(n)$. In contrast, for parameter $log_{10}(K_s)$, the extended data yield a significant improvement, especially for the test case of the FXW model. Specifically, the R^2 increases from 0.22 (original) to 0.56 (extended) and the *RMSE* reduces from 0.87 to 0.68 cm d^{-1}. It should be noted that K_s in this study was treated as a free fitting parameter. However, if the observations of conductivity only cover a limited potential range, the fitted K_s would be biased. The FXW-M model also improves the prediction of $log_{10}(K_s)$ with the extended data, but in a magnitude that is much less than for the FXW model.

The comparison between models H2 and H3 shows that model H3, with the additional input of the bulk density, achieves an overall improvement in predicting the five model parameters.

Table 2. Model H2. The R^2, *RMSE*, and *ME* between the fitted parameters and the predicted ones with the developed PTFs, including both the training and test cases.

Model H2		R^2		*RMSE*		ME	
		EXT	Original	EXT	Original	EXT	Original
$log_{10}(\ )$	Train					−0.0003	−0.0029
	Test						
log_{10}(n)	Train						−0.0005
	Test					−0.0405	−0.0222
log_{10}(m)	Train					−0.0012	
	Test						−0.0062
$_s$	Train						−0.0001
	Test					−0.0022	−0.0039
$log_{10}(K_s)$	Train					−0.0072	
- FXW							
	Test						
$log_{10}(K_s)$	Train					−0.0037	−0.0008
- FXW-M							
	Test					−0.0001	−0.0096

Table 3. Model H3. The R^2, *RMSE*, and *ME* between the fitted parameters and the predicted ones with the developed PTFs, including both the training and test cases.

Model H3		R^2		*RMSE*		ME	
		EXT	Original	EXT	Original	EXT	Original
$log_{10}(\ )$	Train						−0.0012
	Test						
log_{10}(n)	Train					−0.0006	
	Test					−0.0368	−0.0169
log_{10}(m)	Train					−0.0012	−0.0010
	Test						−0.0092
$_s$	Train					−0.0001	
	Test					−0.0036	−0.0044
$log_{10}(K_s)$	Train					−0.0095	−0.0035
- FXW							
	Test						
$log_{10}(K_s)$	Train					−0.0010	
- FXW-M							

Model H3	R^2	*RMSE*	ME
Test			

3.3. Performance of the PTFs

3.3.1. Model H1

For the FXW-M model, the performance of the PTFs developed with the input of the soil texture classes and the extended data is shown in Figure 5. A generally good agreement with the observations is achieved for both the SWRC and HCC, with *RMSE* values of 0.073 cm^3 cm^{-3} and 0.87 cm d^{-1}, and the R^2 being 0.76 and 0.81, respectively.

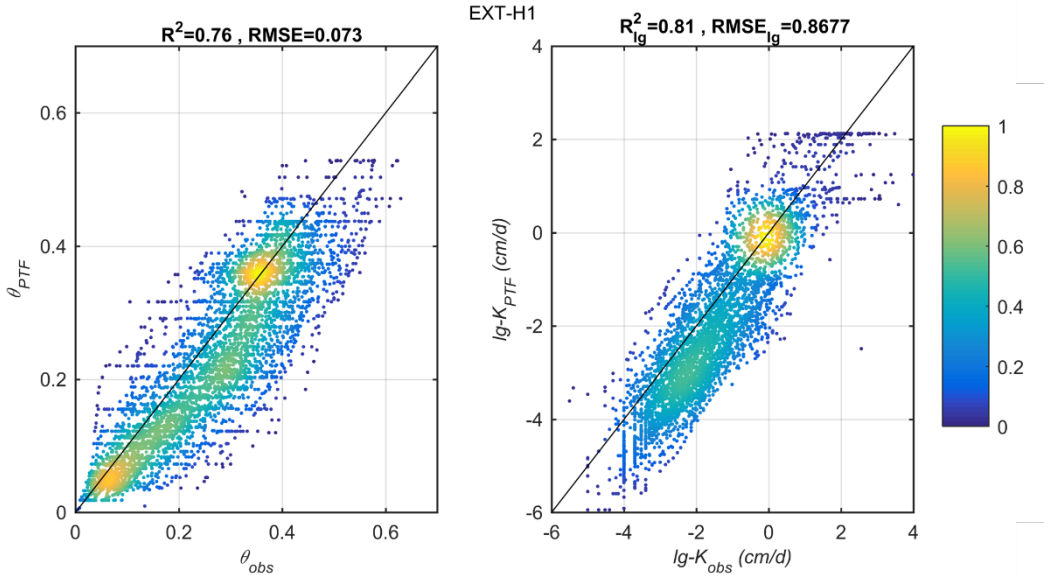

Figure 5. Model H1 performance. The color bar indicates the data density, which is also applied for the other figures.

3.3.2. Prediction of SHPs with the PTFs developed from the original UNSODA data

Figures 6 and 7 present the prediction of SHPs with the different PTFs developed with the original UNSODA data. In the main text, only model H3 with the input of the sand, silt, and clay percentages, as well as the bulk density, is described, while the performance of model H2 is provided in the supplementary material (Figure S3). Three different soil hydraulic models were applied, i.e., the VGM model that accounts for the capillary forces, the FXW model that considers the additional impact of adsorption forces and vapor diffusion, and the modified

FXW-M model that overcomes the abrupt drop near saturation found with the FXW model for soils with n values close to 1.

For the SWRC prediction, the VGM model tends to overestimate the water content in the dry moisture range, while showing underestimation in the wet moisture range (Figure 6). The overall R^2 is 0.87 and the $RMSE$ is 0.050 cm^3 cm^{-3}. The FXW model, in contrast, significantly improves the prediction of water content in both the dry and wet moisture ranges. The reported R^2 is 0.89 and the $RMSE$ is 0.043 cm^3 cm^{-3} when evaluating with all the data. For the test case, the improvement is even more significant. The R^2 increases from 0.83 to 0.87 and the $RMSE$ reduces from 0.054 to 0.045 cm^3 cm^{-3}. It should be noted that both the FXW and FXW-M models have the same SWRC equation.

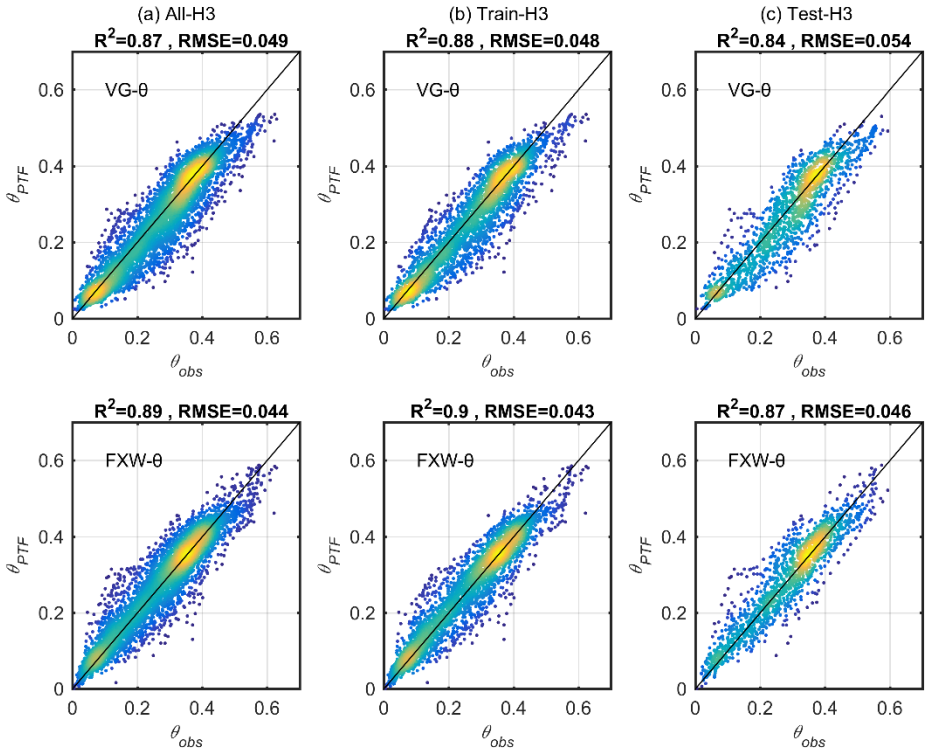

Figure 6. Prediction of the SWRC with the PTFs developed with the original UNSODA data

When it comes to the prediction of the HCC, the PTF developed with the VGM model shows the worst performance, with the R^2 being 0.75 and the $RMSE_{log10(K)}$ being 0.98 cm d^{-1} with the input of the sand, silt, and clay percentages, and the bulk density (model H3). Significant underestimations are apparent for conductivity observations of less than about 0.01 cm d^{-1}. When applying the FXW model that accounts for capillary and adsorption forces as

well as vapor diffusion, the developed PTF improves the prediction in especially dry conditions, yielded a higher R^2 of 0.77 and a lower $RMSE_{log10(K)}$ of 0.83 cm d^{-1}. For the test case, however, the PTF developed with the FXW model overestimates the conductivity, with the $RMSE_{log10(K)}$ being 0.92 cm d^{-1}, which is even higher than that predicted with the VGM model.

Compared to the FXW model, the FXW-M model provided in Wang et al. (2021) overcomes the shortcoming of the abrupt drop near saturation for soils with n values close to 1. With this modified FXW-M model, the developed PTFs significantly improve the predictions. The reported R^2 increases from 0.77 (FXW model) to 0.82, and the $RMSE_{log10(K)}$ decreases from 0.83 (FXW model) to 0.67 cm d^{-1}. Specifically, the FXW-M model greatly improves the prediction of the conductivity in the test case, yielding no overestimation of conductivity. The $RMSE_{log10(K)}$ is 0.72 cm d^{-1}, which is much smaller than the 0.92 cm d^{-1} of the FXW model.

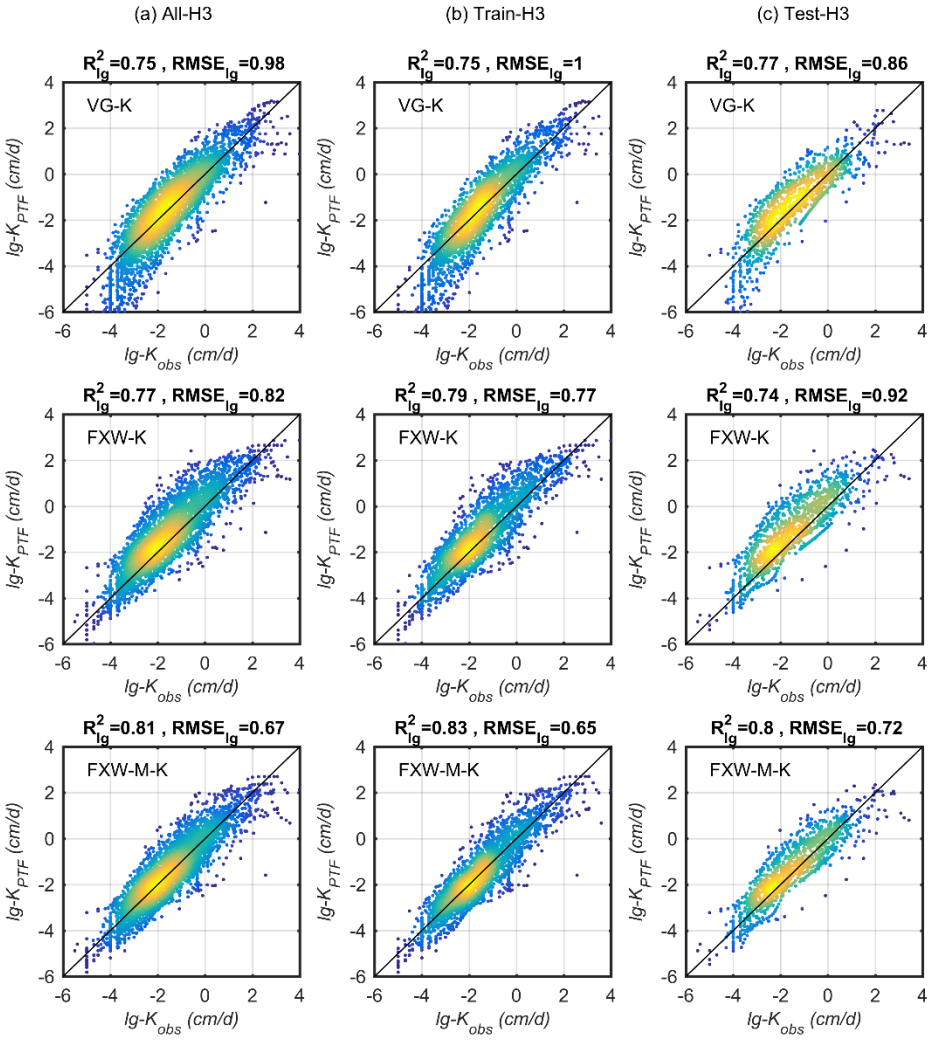

Figure 7. Prediction of the HCC with the PTFs developed with the original UNSODA data

Figure S3 also provides the predictions of conductivity obtained with the PTFs developed with the input of the sand, silt, and clay percentages (model H2). Compared to model H3, the prediction of the water content and hydraulic conductivity with model H2 shows a slightly worse performance for all three soil hydraulic models. Taking the PTFs developed with the FXW-M model as an example, the R^2 decreases from 0.89 to 0.86 and the $RMSE$ increases from 0.044 (model H3) to 0.049 cm^3 cm^{-3} (model H2) when predicting the water content. A close examination of the SWRC prediction shows that the difference occurs

mainly within the range of water content from about 0.4 to 0.6 cm^3 cm^{-3}, i.e., close to the saturated condition, where model H2 shows significant underestimation. For the HCC, model H2 also obtains a lower R^2 of 0.80 and a higher $RMSE_{log10(K)}$ of 0.70 cm d^{-1}, compared to the 0.82 and 0.67 cm d^{-1} of model H3.

3.3.3. Prediction of SHPs with the PTFs developed with the extended UNSODA data

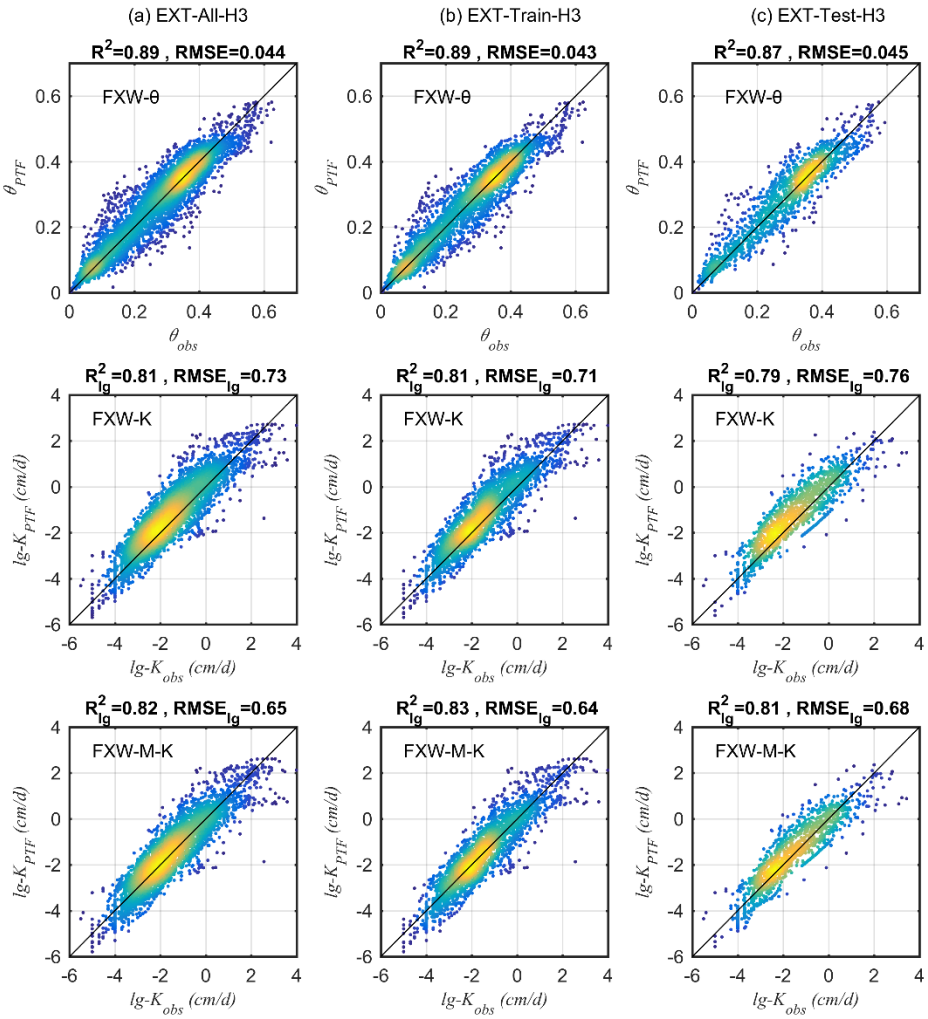

Figure 8. The performance of model H3 with the PTFs developed with the extended UNSODA data

With the extended UNSODA data, the developed PTFs show a significant improvement in predicting the HCC, as shown in Figure 8.

For the FXW model, the PTFs developed with the extended data for model H3 reduce the $RMSE_{log10(K)}$ from 0.82 for the original data to 0.73 cm d^{-1} when evaluating with all the data. When it comes to the test case, the reduction is even more significant, from 0.94 to 0.76 cm d^{-1}.

The PTFs developed with the FXW-M model also yield improved prediction of SHPs with the extended data. Compared to the PTFs developed with the original data, the PTFs developed with the extended data reduce the $RMSE_{log10(K)}$ from 0.67 to 0.65 cm d^{-1} when testing with all the data. The improvement is more significant for the test data, with the $RMSE_{log10(K)}$ reduced from 0.72 to 0.68 cm d^{-1} and the R^2 increased from 0.80 to 0.81.

When it comes to the prediction of the SWRC, however, the PTFs developed with the extended data yield almost the same performance as those developed with the original data.

The PTFs developed with model H2 yield a similar improvement as those with model H3 as shown in Figure S4.

3.4. Extending the capillary-based PTFs to predict SHPs from saturation to oven dryness

With the extended SWRC and HCC in dry conditions, the existing PTFs developed with the capillary-based soil hydraulic models can be easily extended to the entire moisture range. An example applying the Rosetta3 PTF proposed by Zhang and Schaap (2017) is shown in Figure 9.

The Rosetta3 PTF was developed with the capillary-based VGM soil hydraulic model. Due to it not considering the impact of adsorption forces, the VGM model fails to describe the SHPs under dry conditions (Wang et al., 2018). Figure 9 demonstrates that the original Rosetta3 PTF results in obvious overestimation of the soil moisture and underestimation of the hydraulic conductivity in dry conditions. Meanwhile, for most of the data in the medium potential range, the Rosetta3 PTF overestimates the conductivity. The overall R^2 is 0.81 and 0.60, and the $RMSE$ is 0.058 cm^3 cm^{-3} and 1.428 cm d^{-1} for and $log_{10}(K)$, respectively.

By applying the extended method, as described in Section 2.5. the extended Rosetta3 PTF improves the prediction of soil moisture in dry conditions, with the $RMSE$ reduced from 0.058 to 0.056 cm^3 cm^{-3}. The improvement is even more significant for conductivity, with the R^2 increased from 0.60 to 0.66 and the $RMSE_{log10(K)}$ decreased from 1.428 to 1.196 cm d^{-1}.

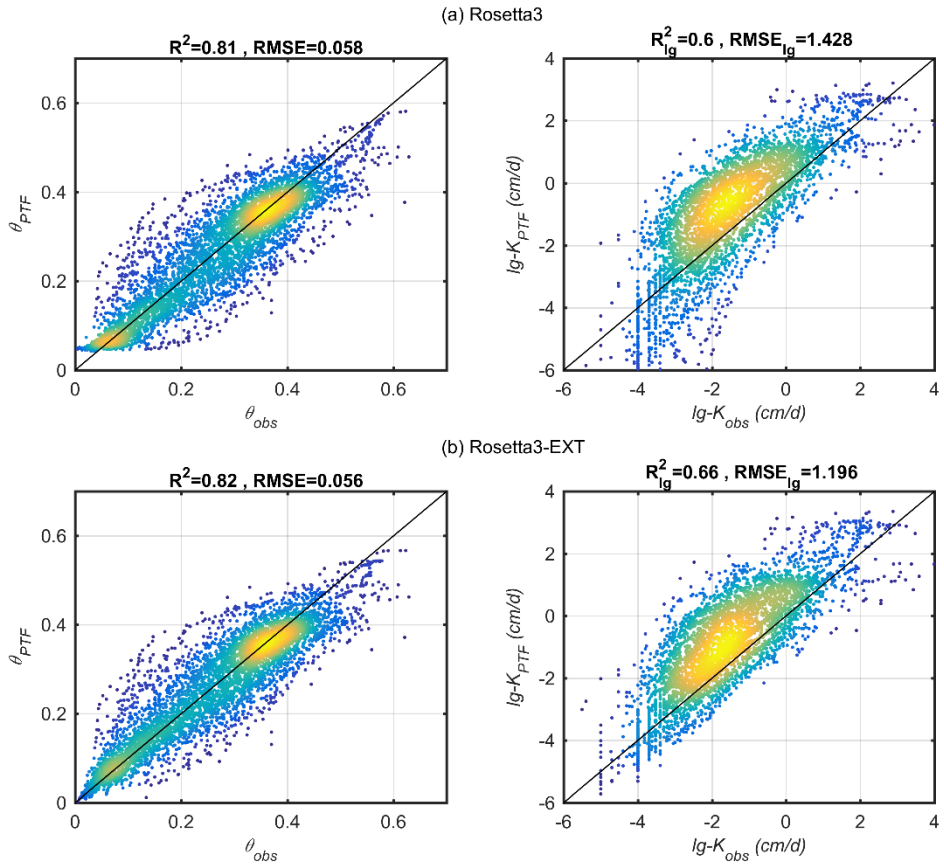

Figure 9. Model performance for the original and the extended Rosetta3 PTF developed with the capillary-based VGM model.

4. Discussion

4.1. Predicting complete SHPs with limited observations

SHPs are determined by adsorption forces under dry conditions. As the specific soil surface area is the key factor controlling the film flow that is held by adsorption forces, this enables us to develop a simple way for predicting SHPs under dry conditions with the known specific soil surface area, which is in turn controlled by easily measured soil texture information, such as the clay fraction and bulk density. When combined with SHPs measured in the capillary-dominant zone, the developed method shows an excellent performance in predicting complete SHPs.

An early work by Jensen et al. (2015) presented a method for predicting the SWRC under dry conditions by building a relationship between a specific water

content (at the potential of -1.0×10^6 cm) with soil texture information through simple linear regression. Only 21 soil samples were applied in this study by Jensen et al. (2015). Here, by applying a more powerful machining learning method (the RF method), a direct relationship between SL and soil texture information was derived based on a total of 275 soil samples. Furthermore, a physically based method was developed for predicting the HCC under dry conditions with the estimated SWRC.

In practice, this method provides a simple and accurate way for deriving complete SHPs with measurements covering only the wet moisture range; for instance, for a matric potential of higher than -1.0×10^3 cm. This matric range is covered and can be easily measured with the widely applied tensiometers. Notably, the exact boundary where capillarity dominates can be determined from equation (13) for a given soil (Lehmann et al., 2008).

In the literature, to derive a complete SWRC, several different devices have often been required. For example, tensiometers for the high matric potential range (0 to -100 KPa), pressure plate apparatus for high to medium dry conditions (0 to several MPa), and chilled-mirror dew point devices (WP4-T for example) for very dry conditions (several to hundreds of MPa). Meanwhile, a long period is required to reach equilibrium under very dry conditions (Wang et al., 2013). Therefore, it is very difficult and time-consuming to measure a complete SWRC directly. For the HCC, measurements in very dry conditions are even more difficult because of the extremely slow water movement rate, and have rarely been presented in the literature.

Accordingly, the method described in this paper represents a simple and accurate method for deriving SHPs from saturation to oven dryness with only measurements from a relatively wet moisture range.

4.2. Limitations of the model structure and SHP observations in PTF development

Most PTFs provided in the literature show a relatively poor performance under low moisture conditions. Taking the Rosetta3 PTF developed with the most commonly applied VGM model as an example, it generally presents overestimation of moisture and obvious underestimation of conductivity under dry conditions (Figure 9, and also reported by Zhang & Schaap, 2017 and Rudiyanto et al., 2021), due to the limitation of the model structure. That is, this kind of model only accounts for capillary forces while neglecting the impact of adsorption forces that are dominant under dry conditions (e.g., Tuller & Or, 2001; Tokunaga, 2009; Wang et al., 2016; 2018). When evaluating with the selected 212 soil samples, the VGM model based Rosetta3 PTF also yields significant overestimation of conductivity for most data in the medium potential range (Figure 9). In Rosetta3, the PTF for predicting the parameter K_s was developed with the measured data. As Schaap and Leij (2000) and Schaap et al. (2001), among many others, have pointed out that applying observed K_s as matching point can lead to overprediction of conductivities at most matric potentials, due

to the K_s is sensitive to macropore flow while unsaturated flow occurs in the soil matrix (van Genuchten & Nielsen, 1985). In contrast, the PTF developed with the fitted K_s improved the predictions of HCC (Figure 7).

When applying the FXW model that considers both capillary and adsorption forces, the developed PTFs improve the prediction of soil moisture and conductivity under dry conditions. However, the developed PTFs with the original data overestimate the conductivity in the test case (Figure 7). This overestimation can be attributed to the abrupt drop near saturation of the FXW model for soils with small n values (Wang et al., 2018; 2021; de Rooij et al., 2021). This shortcoming of dramatic decrease is a result of the non-zero $d\theta/dh$ at the matric potential of zero (van Genuchten & Nielsen, 1985; Schaap & van Genuchten, 2006; de Rooij et al., 2021). For these soils, a much higher K_s is therefore expected when fitting with the observations. Accordingly, if the applied data include a relatively high proportion of soils with small n values, the developed PTF tends to overestimate K_s and then overestimate the unsaturated conductivity for soils with high n values.

In contrast, the PTFs developed with the FXW-M model, which solves the unrealistic drop by introducing a non-zero air-entry value (Wang et al., 2021), significantly improves the prediction of conductivity (Figures 7 and 8). We can also include the impact of vapor diffusion, which only slightly improves the model performance, mainly due to the limited observations of conductivity in extremely dry conditions. The significantly improved performance obtained when considering the impact of adsorption forces and dealing with the unrealistic decrease near saturation indicates that the limitations coming from the model structure have to be considered in PTF development, in addition to applying broader datasets and more model inputs, as well as more powerful machining learning or deep learning methods.

Furthermore, most measurements of SHPs only cover a relatively high matric potential range, which is especially true for the HCC. This represents a great limitation, but is often not considered or dealt with (e.g., Rudiyanto et al., 2021) when developing PTFs that aim to predict SHPs over the entire moisture range. Here, we showed that, by applying extended data that cover the dry conditions, the derived PTFs show a significant improvement in HCC prediction, in particular, compared with those PTFs developed with the original data. This suggests that PTFs developed with limited observations can results in obvious bias in the SHP prediction. For the SWRC, the improvement seen with extended data is not so obvious. This might be due to the difference in training datasets. That is, for the SWRC, about a half of the selected data have observations that cover the potential range of about -1.0×10^4 cm, and might be sufficient for PTF training. When it comes to the HCC, only 29 of the 215 soil samples have measurements for a potential of less than -1.0×10^4 cm. Accordingly, the trained PTF is biased when predicting the HCC in dry conditions. Moreover, one should keep in mind that the lack of measurements in dry conditions also hinders the complete evaluation of the proposed PTFs. With more observations

covering dry conditions, the more physically based PTFs can be expected to show a better performance.

For PTF development with different inputs, the developed PTF with the input of bulk density outperforms that with only the input of soil texture percentages, which is consistent with the findings in other studies of PTF development (e.g., Zhang & Schaap, 2017; Rudiyanto et al., 2021). The main improvements are for prediction in the high water content range, where the H2 model shows obvious underestimation. As confirmed in Tables 2 and 3, model H3 shows an obvious improvement in predicting parameter $_s$ when compared to model H2. This can be attributed to the impact of organic matter and/or the model structure, which can be reflected (partially) by the bulk density (e.g., Minasny & McBratney, 2018; Rawls et al., 2003).

Furthermore, by applying the extended method, we also showed that the existing PTFs developed with the capillary-based soil hydraulic models can be easily extended to achieve an improved performance in the dry moisture range.

In summary, the findings of this study suggest that the impact of the model structure and limited observations has to be considered in PTF development. However, to further improve the PTFs that predict SHPs over the entire moisture range, a much broader dataset and the impact of the soil structure (such as bimodal effects) need to be considered.

5. Concluding Remarks

In this paper, we have presented a simple and accurate method for predicting complete SHPs with measurements taken only in a relatively high matric potential range. Testing with a broad dataset showed that the method performs very well in describing the SWRC (with 213 soil samples) and reasonably well in matching the HCC (with 65 soil samples). This method will be of great importance in practice, considering the difficult and time-consuming nature of measuring SHPs under dry conditions.

Based on this method, the SHPs of 422 soil samples (including 215 samples with HCC measurements) selected from the UNSODA database (Nemes et al., 2001) were extended to the complete moisture range. These data were then further applied in developing and testing the PTFs with the use of different models. The results indicated that the FXW-M model, which accounts for both capillary and adsorption forces and overcomes the unrealistic decrease of the HCC near saturation for fine-textured soils, together with the extended data, obtained the best prediction of SHPs from saturation to oven dryness. This suggests that the model structure and limited observations play an important role in PTF development.

The extended method was further applied together with the Rosetta3 PTF presented by Zhang and Schaap (2017), which accounts for only capillary forces, to predict the SHPs from saturation to oven dryness. The test results yielded an obvious improvement in describing SHPs under dry conditions.

The method presented in this paper and the developed PTFs will be beneficial for the study of soil water flow and the associated processes in relatively dry conditions.

Appendix A.

The van Genuchten (1980)-Mualem (1976) model (known as the VGM model) is denoted as:

and

where S is the efficient water saturation degree in relation to capillary water; $_s$ and $_r$ are the saturated water content and the residual water content, respectively; K (L T^{-1}) is the soil hydraulic conductivity; K_s (L T^{-1}) is the saturated hydraulic conductivity; and (L^{-1}), n, m=1−1/n, l (which generally have a value of 0.5) are all fitting parameters.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (Nos. 42071045, 41722208), in part by the Fundamental Research Funds for Central Universities, in part by the China University of Geosciences (Wuhan) (No. CUG1323531877), and in part by the Natural Sciences Foundation of Hubei Province of China (2019CFA013). The applied data were obtained from a a public dataset, which is available at the website of the United States Department of Agriculture (https://data.nal.usda.gov/search/type/dataset).

References

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.

Arthur, E., Tuller, M., Moldrup, P., Resurreccion, A. C., Meding, M. S., Kawamoto, K., ... & De Jonge, L. W. (2013). Soil specific surface area and non-singularity of soil-water retention at low saturations. *Soil Science Society of America Journal*, *77*(1), 43-53.

Bird, R. B., W. E. Stewart, and E. N. Lightfoot (1960), Transport Phenomena, John Wiley, New York.

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. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(6), 493-507.

Børgesen, C. D., & Schaap, M. G. (2005). Point and parameter pedotransfer functions for water retention predictions for Danish soils. *Geoderma*, *127*(1-2),

154-167.

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: An International Journal*, *22*(11), 1630-1639.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Chen, C., Hu, K., Ren, T., Liang, Y., & Arthur, E. (2017). A simple method for determining the critical point of the soil water retention curve. *Soil Science Society of America Journal*, *81*(2), 250-258.

Campbell, G.S., Shiozawa, S., (1992). Prediction of hydraulic properties of soils using particle-size distribution and bulk density data. In: van Genuchten, M.Th., Leij, F.J., Lund, L.J. (Eds.), *Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soil*. Univ. of California, Riverside, CA, USA, pp. 317–328.

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783-2792.

Dai, Y., Xin, Q., Wei, N., Zhang, Y., Shangguan, W., Yuan, H., ... & Lu, X. (2019). A global high-resolution data set of soil hydraulic and thermal properties for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *11*(9), 2996-3023.

de Rooij, G. H., Mai, J., & Madi, R. (2021). Sigmoidal water retention function with improved behaviour in dry and wet soils. *Hydrology and Earth System Sciences*, *25*(2), 983-1007.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, *28*(4), 1015-1031.

Fayer, M. J., & Simmons, C. S. (1995). Modified soil water retention functions for all matric suctions. *Water Resources Research*, *31*(5), 1233-1238.

Fredlund, D. G., Xing, A., & Huang, S. (1994). Predicting the permeability function for unsaturated soils using the soil-water characteristic curve. *Canadian Geotechnical Journal*, *31*(4), 533–546. https://doi.org/10.1139/t94-062

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, *12*(2), e0169748.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518.

Ippisch, O., Vogel, H. J., & Bastian, P. (2006). Validity limits for the van Genuchten–Mualem model and implications for parameter estimation and numerical simulation. *Advances in water resources, 29*(12), 1780-1789.

Iwamatsu, M., & Horii, K. (1996). Capillary condensation and adhesion of two wetter surfaces. *Journal of colloid and interface science, 182*(2), 400-406.

Jensen, D. K., Tuller, M., de Jonge, L. W., Arthur, E., & Moldrup, P. (2015). A new two-stage approach to predicting the soil water characteristic from saturation to oven-dryness. *Journal of Hydrology*, *521*, 498-507.

Lamorski, K., Pachepsky, Y., Sławiński, C., & Walczak, R. T. (2008). Using support vector machines to develop pedotransfer functions for water retention of soils in Poland.

Langmuir, I. (1938). Repulsive forces between charged surfaces in water, and the cause of the Jones-Ray effect. *Science*, *88*(2288), 430-432.

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

Lehmann, P., Assouline, S., & Or, D. (2008). Characteristic lengths affecting evaporative drying of porous media. *Physical Review E*, *77*(5), 056309.

Liao, K., Lai, X., Zhou, Z., Zhu, Q., & Han, Q. (2018). A simple and improved model for describing soil hydraulic properties from saturation to oven dryness. *Vadose Zone Journal*, *17*(1), 1-8.

Lu, S., Ren, T., Gong, Y., & Horton, R. (2008). Evaluation of three models that describe soil water retention curves from saturation to oven dryness. *Soil Science Society of America Journal, 72*(6), 1542-1546.

Lu, S., Ren, T., Lu, Y., Meng, P., & Sun, S. (2014). Extrapolative capability of two models that estimating soil water retention curve between saturation and oven dryness. *PloS one,* 9(12), e113518.

Millington, R. J., & Quirk, J. P. (1961). Permeability of porous solids. *Transactions of the Faraday Society*, 57, 1200-1207.

Minasny, B., & McBratney, A. B. (2018). Limited effect of organic matter on soil available water capacity. *European journal of soil science*, *69*(1), 39-47.

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

Nemes, A. D., 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.

Nemes, A., Rawls, W. J., & Pachepsky, Y. A. (2006). Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Science Society of America Journal*, *70*(2), 327-336.

Nimmo, J. R. (1991). Comment on the treatment of residual water content in "A consistent set of parametric models for the two-phase flow of immiscible fluids in the subsurface" by L. Luckner et al. *Water Resources Research*, *27*(4), 661-662.

Or, D., & Tuller, M. (2000). Flow in unsaturated fractured porous media: Hydraulic conductivity of rough surfaces. *Water Resources Research*, *36*(5), 1165-1177.

Pachepsky, Y., & Rawls, W. J. (Eds.). (2004). *Development of pedotransfer functions in soil hydrology* (Vol. 30). Elsevier.

Pachepsky, Y. A., Rawls, W. J., & Lin, H. S. (2006). Hydropedology and pedotransfer functions. *Geoderma*, *131*(3-4), 308-316.

Rawls, W. J., & Pachepsky, Y. A. (2002). Using field topographic descriptors to estimate soil water retention. *Soil Science*, *167*(7), 423-435.

Rawls, W. J., Pachepsky, Y. A., Ritchie, J. C., Sobecki, T. M., & Bloodworth, H. (2003). Effect of soil organic carbon on soil water retention. *Geoderma*, *116*(1-2), 61-76.

Resurreccion, A. C., Moldrup, P., Tuller, M., Ferré, T. P. A., Kawamoto, K., Komatsu, T., & De Jonge, L. W. (2011). Relationship between specific surface area and the dry end of the water retention curve for soils with varying clay and organic carbon contents. *Water Resources Research, 47*(6).

Robinson, A. (2015). Environmental science: Dryland epiphany. *Nature*, *518*(7540), 482-482.

Rossi, C., & Nimmo, J. R. (1994). Modeling of soil water retention from saturation to oven dryness. *Water Resources Research*, *30*(3), 701-708.

Rudiyanto., Minasny, B., Chaney, N. W., Maggi, F., Giap, S. G. E., Shah, R. M., Fiantis, D., & Setiawan, B. I. (2021). Pedotransfer functions for estimating soil hydraulic properties from saturation to dryness. *Geoderma*, *403*, 115194.

Saito, H., Šimůnek, J., & Mohanty, B. P. (2006). Numerical analysis of coupled water, vapor, and heat transport in the vadose zone. *Vadose Zone Journal, 5*(2), 784–800.

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*(3), 843-851.

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.

Schneider, M., & Goss, K. U. (2012). Prediction of the water sorption isotherm in air dry soils. *Geoderma*, *170*, 64-69.

Stanić, F., Delage, P., Tchiguirinskaia, I., Versini, P. A., Cui, Y. J., & Schertzer, D. (2020). A new fractal approach to account for capillary and adsorption phenomena in the water retention and transfer properties of unsaturated soils. *Water Resources Research*, *56*(12), e2020WR027808.

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*(1), 151-175.

Tokunaga, T. K. (2009), Hydraulic properties of adsorbed water films in unsaturated porous media, *Water Resour. Res.*, *45*, W06415, doi:10.1029/2009WR007734.

Torres, E. A., & Calera, A. (2010). Bare soil evaporation under high evaporation demand: a proposed modification to the FAO-56 model. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques, 55*(3), 303-315.

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*(1), 226-238.

Tsujimura, M., Numaguti, A., Tian, L., Hashimoto, S., Sugimoto, A., & Nakawo, M. (2001). Behavior of subsurface water revealed by stable isotope and tensiometric observation in the Tibetan Plateau. *Journal of the Meteorological Society of Japan. Ser. II, 79*(1B), 599-605.

Tuller, M., and D. Or (2001), Hydraulic conductivity of variably saturated porous media: Film and corner flow in angular pore space, *Water Resour. Res.*, *37(5)*, 1257–1276, doi:10.1029/2000WR900328.

Tuller, M., and D. Or (2005), Water films and scaling of soil characteristic curves at low water contents, *Water Resources Research*, *41*(9).

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.

van Genuchten, M. T., & Nielsen, D. R. (1985). On describing and predicting the hydraulic properties. In *Annales Geophysicae* (Vol. 3, No. 5, pp. 615-628).

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., ... & Vereecken, H. (2017). Pedotransfer functions in Earth system science: challenges and perspectives. *Reviews of Geophysics*, *55*(4), 1199-1256.

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M. G., & Genuchten, M. T. V. (2010). Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: A review. *Vadose Zone Journal*, *9*(4), 795-820.

Vogel, T., Van Genuchten, M. T., & Cislerova, M. (2000). Effect of the shape of the soil hydraulic functions near saturation on variably-saturated flow predictions. *Advances in water resources*, *24*(2), 133-144.

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*(3-4), 169-185.

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*(3-4), 123-150.

Wright MN, Ziegler A (2017). "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software*, 77(1), 1–17.

Wang, Y., Ma, J., Zhang, Y., Zhao, M., & Edmunds, W. M. (2013). A new theoretical model accounting for film flow in unsaturated porous media. *Water Resources Research*, *49*(8), 5021-5028.

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.

Wang, Y., Ma, J., Guan, H., & Zhu, G. (2017). Determination of the saturated film conductivity to improve the EMFX model in describing the soil hydraulic properties over the entire moisture range. *Journal of hydrology*, *549*, 38-49.

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.

Wang, Y., Merlin, O., Zhu, G., & Zhang, K. (2019). A physically based method for soil evaporation estimation by revisiting the soil drying process. *Water Resources Research*, 55, 9092– 9110.

Wang, Y., Ma, R., & Zhu, G. (2021). Improved Prediction of Hydraulic Conductivity with Soil Water Retention Curve that Accounts for Both Capillary and Adsorption Forces. *Earth and Space Science Open Archive*, 10.1002/essoar.10508079.1. https://doi.org/10.1002/essoar.10508079.1.

Webb, S. W. (2000). A simple extension of two-phase characteristic curves to include the dry region. *Water Resources Research, 36*(6), 1425-1430.

Weber, T. K., 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.

Weynants, M., Montanarella, L., Toth, G., Arnoldussen, A., Anaya Romero, M., Bilas, G., ... & Woesten, H. (2013). European HYdropedological Data Inventory (EU-HYDI). *EUR Scientific and Technical Research Series*.

Zhang, Z. F. (2011). Soil water retention and relative permeability for conditions from oven-dry to full saturation. *Vadose Zone Journal*, *10*(4), 1299-1308.

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.

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.