water

MDPI

Article

Metropolis-Hastings Markov Chain Monte Carlo Approach to Simulate van Genuchten Model Parameters for Soil Water Retention Curve

Xuan Du [1,†], Can Du [2,†], Jesse Radolinski [3], Qianfeng Wang [4] and Jinshi Jian [5,6,*]

1 Department of Hydraulic Engineering, Yangling Vocational & Technical College, XianYang 712100, China; lemon_19988@163.com
2 Department of Bioengineering, Yangling Vocational & Technical College, XianYang 712100, China; ducan616@163.com
3 Department of Ecology, University of Innsbruck, 6020 Innsbruck, Austria; Jesse.Radolinski@uibk.ac.at
4 Fujian Provincial Key Laboratory of Remote Sensing of Soil Erosion, College of Environmental & Safety Engineering, Fuzhou University, Fuzhou 350116, China; wangqianfeng@fzu.edu.cn
5 State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A & F University, XianYang 712100, China
6 Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resource, XianYang 712100, China
* Correspondence: jinshi@nwafu.edu.cn; Tel.: +86-182-0290-1951
† These authors contributed equally to this work.

check for updates

Citation: Du, X.; Du, C.; Radolinski, J.; Wang, Q.; Jian, J. Metropolis-Hastings Markov Chain Monte Carlo Approach to Simulate van Genuchten Model Parameters for Soil Water Retention Curve. *Water* **2022**, *14*, 1968. https://doi.org/10.3390/w14121968

Academic Editor: Mirko Castellini

Received: 12 May 2022
Accepted: 18 June 2022
Published: 20 June 2022

Abstract: The soil water retention curve (SWRC) is essential for assessing water flow and solute transport in unsaturated media. The van Genuchten (VG) model is widely used to describe the SWRC; however, estimation of its effective hydraulic parameters is often prone to error, especially when data exist for only a limited range of matric potential. We developed a Metropolis-Hastings algorithm of the Markov chain Monte Carlo (MH-MCMC) approach using R to estimate VG parameters, which produces a numerical estimate of the joint posterior distribution of model parameters, including fully-quantified uncertainties. When VG model parameters were obtained using complete range of soil water content (SWC) data (i.e., from saturation to oven dryness), the MH-MCMC approach returned similar accuracy as the widely used non-linear curve-fitting program RETC (RETention Curve), but avoiding non-convergence issues. When VG model parameters were obtained using 5 SWC data measured at matric potential of around −60, −100, −200, −500, and −15,000 cm, the MH-MCMC approach was more robust than the RETC program. The performance of MH-MCMC are generally good ($R^2 > 0.95$) for all 8 soils, whereas the RETC underperformed for coarse-textured soils. The MH-MCMC approach was used to obtain VG model parameters for all 1871 soils in the National Cooperative Soil Characterization dataset with SWC measured at matric potentials of −60 cm, −100 cm, −330 cm, and −15,000 cm; the results showed that the simulated SWC by MH-MCMC model were highly consistent with the measured SWC at corresponding matric potential. Altogether, our new MH-MCMC approach to solving the VG model is more robust to limited coverage of soil matric potential when compared to the RETC procedures, making it an effective alternative to traditional water retention solvers. We developed an MH-MCMC code in R for solving VG model parameters, which can be found at the GitHub repository.

Keywords: soil water retention curve; van Genuchten; Bayes; Markov Chain Monte Carlo

1. Introduction

The relationship between volumetric soil water content (SWC, θ, $L^3 \cdot L^{-3}$) and matric potential (*h*, -cm, taken negative for increasing suctions, same hereafter), i.e., the soil water retention curve (SWRC), is critical for understanding the hydraulic characteristics of a soil. The SWRC is a physical basis for deriving flow and storage properties of variably saturated porous media and, as such, is key to quantifying water (e.g., infiltration, evaporation, and root and water uptake) and material (e.g., erosion, weathering, and solute transport) fluxes in soils. Therefore, it is important to develop methods which can accurately describe the SWRC for different soils.

The SWRC is typically determined under laboratory conditions where h or SWC are measured and used to build an empirical retention curve. To date, no one measurement device allows for full determination of water retention over all possible moisture contents; thus it has become common to use multiple measurement methods to build a complete SWRC [1]. For example, the SWRC is often determined using a porous plate apparatus, in which SWC can be measured gravimetrically for discrete h values after pressure equilibration with the ceramic plate matrix [2,3]. However, because ceramic plate methods have lower accuracy in the driest range (e.g., $h < -4000$ cm) [4], they can be supplemented with methods that measure water potential in the vapor phase to derive h (e.g., dew point methods) [5]. These values become the basis for extrapolating an accurate SWRC for a given soil.

Many mathematical models have been developed for estimating the SWRC of soils. For example: Brooks and Corey [6], van Genuchten [7], Fayer and Simmons [8], Webb [9], Khlosi et al. [10], Fredlund and Xing [11], and Groenevelt and Grant [12] developed models to fit the SWRC data from saturation to oven dryness. Among these models, the van Genuchten model (hereafter called VG model) is most widely used because it well describes the SWRC for a large range of soil types [13–15]. The VG model, however, is a complex non-linear equation, involving several parameters (i.e., saturated water contents (θ_S), residual water contents (θ_r), a parameter related to the inverse of the air entry pressure (α), and a metric related to the pore-size distribution (n)) to simulate the SWRC curves, making it difficult to obtain optimal solutions for those parameters through inverse modeling.

Modern computing capabilities have facilitated the use of several numerical solvers and algorithms which improve empirically-derived solutions to optimize VG parameters. The widely used RETC (RETention Curve) program [13] is a non-linear, least squares optimization algorithm for building soil water retention curves. Despite the success of the RETC, the program often produces compounding uncertainties when seeking optimal parameter solutions [14,15]. For example, when solving for SWRC parameters, Wang et al. [16] suggests that RETC does not guarantee convergence to the global optimum because it requires the prior soil information to initialize the VG model parameters. The traditional numerical methods to solve for VG parameters usually produce uncertainty and error, and make it difficult to obtain global parameter solutions [17]. Non-convergence, returning extreme parameter values, and inefficient computation are also common problems when seeking solutions for the VG model parameters [14,15]. When combined with the inevitable uncertainty from observational water retention data, the RETC program can produce solver errors that are further manifested in optimization, and eventually, the scientific inference from those results.

A popular and promising Bayesian method, the so-called Markov Chain Monte Carlo (MCMC) approach, is now widely used for a variety of inverse problems in applied mathematics [18] and recently in hydrological simulations [19,20]. Thus, MCMC methods are likely to become useful for solving VG parameters. For example, Carsel et al. [21] used a Monte Carlo approach to characterize input parameters for the pesticide root zone model (PRZM), which then simulated the leaching potential of pesticides. Duan and Gupta [22] presented a shuffled complex evolution algorithm (SCE-UA) to consistently locate the global optimum of a conceptual rainfall-runoff model. Subsequently, an MCMC method was combined with the SCE-UA approach, which drastically improved its computational efficiency [23]. Shi et al. [15] used an adaptive Metropolis MCMC to estimate the parameters of VG model for a silt soil from QingDao, China, and found the adaptive Metropolis MCMC approach to be an effective approach in solving VG model parameters, yet this analysis was restricted to one silt soil from QingDao, China.

The MCMC approach has been successfully used to obtain parameters for the VG model; however, it is unclear whether the MCMC approach is accurate and robust when applied to a wide range of soils. Additionally, the performance of the MCMC approach in estimating VG parameters is unclear when only a limited number of measurements are available (e.g., with h values between −15,000 and −60 cm). To explore these questions, the objectives of this study were: (1) to develop an MCMC program using R which can be used to characterize the entire SWRC using measurements from saturation to dryness; (2) to investigate the performance of the developed MCMC program if only data between −15,306 cm and −50 cm were used; (3) to obtain the optimal solutions for the VG model parameters for all 1871 soils from the National Cooperative Soil Characterization (NCSS) dataset.

2. Materials and Methods

2.1. Soil Water Content under Certain Matric Potential Gradient

Soil water content data from saturation to air dry matric potential gradient of 8 soils were used derived from Lu et al. [24,25]. The soil samples were collected from China and USA, covered a wide range of soil texture, from sand to silt loam. SWC between 0 and −15,000 cm matric potential range were obtained using the pressure plate method, and SWC beyond −15,000 cm matric potential were obtained using the Dewpoint Potential Meter (Model WP4-T, Decagon Device, Pullman, WA, USA). For more detailed information about the soil properties and SWC measurements, please refer to Lu et al. [24,25].

Soil water content at matric potential of 0 cm, −60 cm, −100 cm, −330 cm, and −15,000 cm from the NCSS database were used to train and test the MCMC model in this study. The NCSS includes more than 100,000 samples data from the Kellogg Soil Survey Laboratory and cooperating universities, which contains measurements of soil texture, organic carbon content, bulk density, and SWC under specified matric potential from American soils. A Microsoft Access NCSS database can be accessed through https://ncsslabdatamart.sc.egov.usda.gov/ (accessed on 17 June 2022). In addition to commonly requested data, the Access database includes metadata tables that describe the column headings of the laboratory data tables. Only soil samples with SWC at 0 cm, −60 cm, −100 cm, −330 cm, and −15,000 cm were used in this study. The original water content data were measured gravimetrically and were converted to volumetric water content values by the corresponding bulk density. In addition, we also filtered the data for outliers, and excluded data with organic carbon content larger than 10% and soil with andic properties because they behaved differently from mineral soil. After filtering, there were 1871 data points, distributed into 12 United States Department of Agriculture (USDA) soil texture groups, which were used in this study (Figure 1a,b, light-blue points). Sand clay and silt soil samples were relatively scarce, but all other soil textures have extensive data (Figure 1c,d).

Another subset of the unsaturated soil hydraulic properties database (UNSODA) was used in this study to evaluate the performance of our MCMC model. UNSODA contains water retention, hydraulic conductivity, soil water diffusivity, basic soil properties (e.g., particle-size distribution, bulk density, organic matter content), and additional information regarding the soil and the experimental procedures. UNSODA was used to support the Rosetta model [26]. The Rosetta install software package has a testing subset of UNSODA (555 samples, Figure 1, red points); this subset was used in this study to evaluate the performance of our MCMC model.

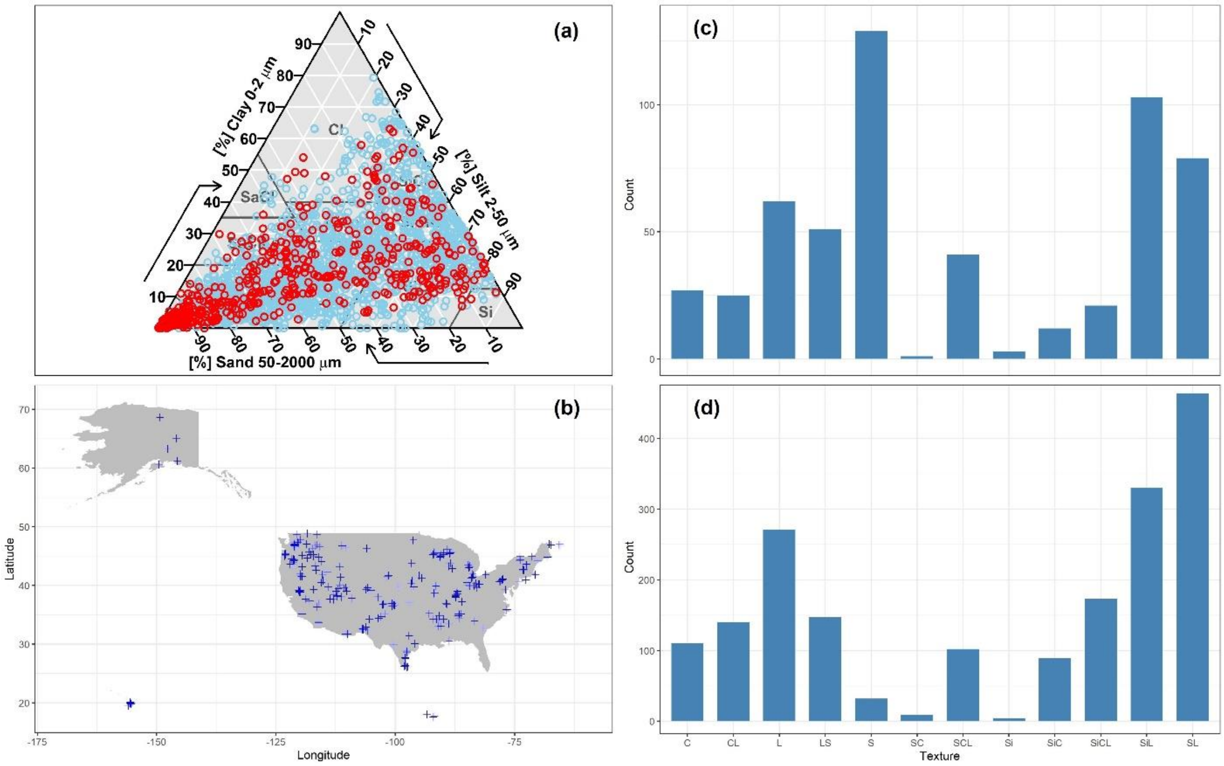

Figure 1. (**a**) Textural distribution of 1871 samples from the National Cooperative Soil Characterization database (NCSS, blue dots) and 555 samples from the Unsaturated Soil Hydraulic Properties Database (UNSODA, red dots) used in this study according to the United States Department of Agriculture (USDA) soil texture system. (**b**) Locations of the samples from NCSS database; note that samples from UNSODA were not shown due to lacking latitude and longitude information. (**c**) Number of soil samples in the UNSODA dataset. (**d**) Number of soil samples in the NCSS dataset. C—clay, CL—clay loam, L—loam, LS—loamy sand, S—sand, SC—sandy clay, SCL—sandy clay loam, Si—silt, SiC—silty clay, SiCL—silty clay loam, SiL—silt loam, SL—sandy loam.

2.2. The van Genuchten (VG) Model

The VG model to describe the SWRC can be explained by Equation (1):

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

where $\theta(h)$ is the measured volumetric water content ($L^3 \cdot L^{-3}$) at the suction h (-cm, taken negative for increasing suctions). The parameters θ_s and θ_r are saturated and residual water contents, respectively ($L^3 \cdot L^{-3}$). α is a positive value (in unit of cm^{-1}), related to the inverse of the air entry pressure, and n (>1, unitless) is a metric related to the pore-size distribution. Both α and n determined the shape of the SWCR [7,13] (Figure A1).

2.3. Markov Chain Monte Carlo (MCMC) Approach

Bayesian methods have two important advantages over traditional model curve-fitting approaches: first, they allow virtually infinite flexibility in deviating from the distributional assumptions of typical statistical methods; second, they provide robust estimates of uncertainty. Practical application of Bayesian methods is challenged by the need to compute high-dimensional integrals associated with the interactions of many probability distribution functions. To a large extent, this has been solved by the advent of Markov Chain Monte Carlo (MCMC) methods, which leverage dramatic recent increases in computing power to estimate these integrals numerically [18,27].

2.4. Obtaining Parameters of VG Model Using the MH-MCMC Approach

When using the MCMC approach to obtain the posterior distribution of the VG model parameters, the model could be written as Equation (2). According to previous studies [15,27–29], we assumed a uniform distribution for the prior distribution of every parameter, and the bounds of the parameters were set up according to the distribution of VG model parameters (θ_s, θ_r, α, and n) from the UNSODA (Table 1 and Figure 2).

$$\theta = f(\theta_s, \theta_r, \alpha, n, h) \tag{2}$$

Table 1. Distributions of the parameters for using the MCMC modeling to simulate the soil water content curve. UNSODA: the unsaturated soil hydraulic properties database.

Parameters	Prior	Note
θ_s	0.30–0.80	Adjusted based on UNSODA (Figure 2) and ref. [15]
θ_r	0.0003–0.30	Adjusted based on UNSODA (Figure 2) and ref. [15]
α	0.0001–1.0	Adjusted based on UNSODA (Figure 2) and ref. [15]
n	1.0–10	Adjusted based on UNSODA (Figure 2) and ref. [15]

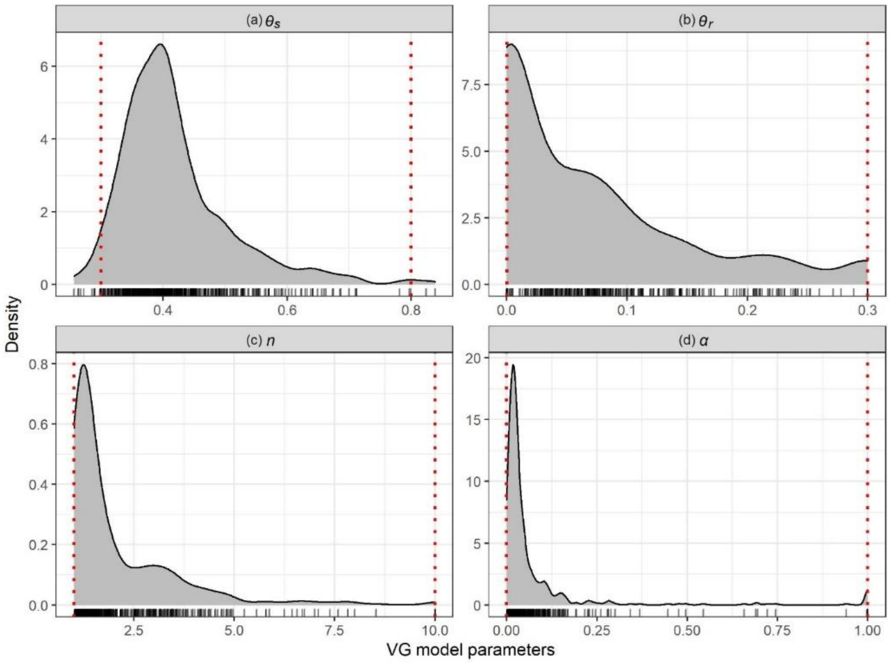

Figure 2. Distribution of (**a**) θ_s, (**b**) θ_r, (**c**) n, and (**d**) α for the van Genuchten water retention curve model parameters from the unsaturated soil hydraulic properties database (UNSODA) predicted by the ROSETTA model [26]. The vertical lines in each panel are low and high bounds of uniform prior used in this study; the gray lines in the x-axis show the data density of each parameter (total n = 554).

The Markov Chain stationary distribution (π) and its transfer matrix (Q) are critical foundations for the MCMC method. Many stochastic simulation methods, including Metropolis–Hastings-MCMC (MH-MCMC) [28,29], adapted MH-MCMC [27], and Gibbs sampling [30] have been developed to resolve this problem. We used the MH-MCMC approach to generate the posterior distribution of all four parameters (θ_s, θ_r, α, and n) of the VG model because MH-MCMC has been successfully used to solve hydrology-related studies. For a detailed description about MH-MCMC, please refer to [28,29]. The sampling process of the MH-MCMC algorithm can be described as follows:

(1) Initiating the model parameters, as MH-MCMC is not sensitive to the initial condition, we therefore set a same values for all soil samples, i.e., set θ_s = 0.56, θ_r = 0.18, α = 0.049, and n = 1.5 as initial values for the VG model parameters.

(2) Based on the model and the measured SWC under different pressure heads to get an initial estimate of θ and h.
(3) Generating an arbitrary Markov Chain stationary distribution $\pi\ (x)$ and its transfer matrix Q. Sample from any simple probability distribution to get the initial state value, x_0.
(4) Set accept rate = $min\left\{\frac{\pi(j)Q(j,i)}{\pi(i)Q(i,j)},\ 1\right\}$.
(5) Sample from the conditional probability distribution $Q(x|x_t)$, get x^*; sample from the uniform distribution $\mu \sim$ U $[0,1]$; if $\mu < \alpha(x_t,\ x^*) = min\left\{\frac{\pi(j)Q(j,i)}{\pi(i)Q(i,j)},\ 1\right\}$, accept $\alpha(x_t,\ x^*)$, i.e., $x_{t+1} = x^*$; and otherwise, reject transformation, i.e., $x_{t+1} = x_t$.

The sample set $(x_{n_1},\ x_{n_1+1},\ \ldots,\ x_{n_1+n_2-1})$ is corresponding to the stationary distribution; repeat this process for all four parameters (θ_s, θ_r, α, and n), and we can get the parameters for the VG equation for each soil sample. We conducted 10,000 iterations of sampling and discarded the first 2000 iterations as the burn-in period. We selected the burn-in amount and evaluated convergence by visual inspection of the MCMC chains. The MH-MCMC program and statistical analyses of the output were conducted using R (Version 4.2.0, R Core Team, 2019). Details on the MH-MCMC algorithm were described in the source code, which are available at the following GitHub repository (https://github.com/jinshijian/SWRC_MCMC, accessed on 17 June 2022).

We used the MH-MCMC approach to obtain VG model parameters for all 8 soil samples in the Lu et al. dataset [24,25], and detailed information for those 8 soil samples can be found in Table A1. We first used MH-MCMC approach to obtain parameters of the VG model (θ_s, θ_r, α, and n) based on all available SWC measurements (15 to 24 SWC measurements from saturation to oven dryness condition, see Table 2). We also used the MH-MCMC approach to obtain VG parameters based on SWC measured at matric potential approximated to −60, −100, −200, −500, and −15,000 cm (Table A1).

Table 2. Summary of adjusted R^2, root mean square error (RMSE), mean error (ME), and number of samples when comparing the van Genuchten (VG) model's predicted and measured soil water content for 8 soils from ref. [24,25].

Soil	*Adjusted* R^2	*RMSE*	*ME*	*Adjusted* R^2	*RMSE*	*ME*	Samples
All SWC Measurements Were Used to Parameterize the VG Model							
	MCMC			**RETC**			
I (Sand)	0.997	0.006	<0.0001	0.998	0.006	0.0007	19
II (Sandy loam)	0.985	0.018	−0.002	0.985	0.018	<0.0001	19
III (Loam)	0.991	0.016	−0.001	0.991	0.016	<0.0001	19
IV (Silt loam)	0.993	0.013	−0.005	0.995	0.012	<0.0001	17
V (Silt clay loam)	0.995	0.010	−0.003	0.996	0.009	0.0002	19
VI (Silt loam)	0.991	0.013	−0.005	0.992	0.012	<0.0001	15
VII (Silt clay loam)	0.992	0.013	−0.006	0.993	0.013	<0.0001	16
VIII (Silt loam)	0.991	0.015	−0.003	0.993	0.013	0.0001	24
Only 5 SWC measurements were used to parameterize the VG model (but all data were used for model performance evaluation)							
	MCMC			**RETC**			
I (Sand)	0.950	0.027	0.013	0.872	0.044	−0.0454	19
II (Sandy loam)	0.956	0.031	0.006	0.900	0.046	0.0182	19
III (Loam)	0.970	0.029	0.007	0.988	0.018	0.0026	19
IV (Silt loam)	0.952	0.037	0.005	0.972	0.028	0.0129	17
V (Silt clay loam)	0.978	0.021	0.002	0.994	0.011	0.0032	19
VI (Silt loam)	0.985	0.017	0.004	0.963	0.026	0.0141	15
VII (Silt clay loam)	0.968	0.027	0.002	0.911	0.045	0.0246	16
VIII (Silt loam)	0.971	0.027	0.018	0.993	0.014	0.0136	24

For the 1871 soils in the NCSS dataset, only SWC at matric potential of 0, −60, −100, −330, and −15,000 cm were available. We also used the MH-MCMC approach to obtain the parameters of VG model using SWC at matric potential of −60, −100, −330, and −15,000 cm. SWC at matric potential of 0 cm (saturated SWC) were not used in the MH-MCMC process but were compared with the θ_s from the MH-MCMC to evaluate its performance.

2.5. Obtaining Parameters of VG Model Using the RETC Program

As a comparison, we also used the RETC program (version 6.02) to determine the best parameter of VG model for the Lu et al. dataset [24,25]. For instruction about RETC usage, please refer to [13]. Same as the MH-MCMC approach, we used the RETC program to characterize the SWRC for 8 soils from Lu et al.'s dataset [24,25] using all measurements from saturation to dryness. As well, we only used 5 measurements at matric potential of around −60, −100, −200, −500, and −15,000 cm (Table A1).

2.6. Model Evaluation

The performance of both the MCMC approach and the RETC program were evaluated by comparing the simulated soil water retention curve and measured data. The distribution, the random walk trace, and the auto-correlation of posteriori estimates of each parameter (θ_s, θ_r, α, and n) were also used to evaluate the performance of the MH-MCMC approach. In addition, the *adj* R^2 (Equation (3)), Root Mean Square Error (RMSE, Equation (4)), and mean error (ME, Equation (5)) were used to evaluate the model performance:

$$adj\ R^2 = 1 - \left[\frac{\left(1 - R^2\right)(n-1)}{n-k-1}\right] \tag{3}$$

where R^2, n, and k are the coefficient of determination, the total number of observations, and the total number of independent variables in the model, respectively. In this study, the *adj* R^2 was used to quantify the variability in the measured $\theta(h)$ as explained by the VG model, which was parameterized using the MH-MCMC approach:

$$ME = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{5}$$

where $\hat{y}_i$ represents the i_{th} predicted $\theta(h)$ value and y_i represents the i_{th} measured $\theta(h)$ value. ME is the averaged difference between predicted and measured $\theta(h)$. Therefore, when $ME \approx 0$, the predicted $\theta(h)$ is not different from the measured $\theta(h)$, whereas $ME > 0$ and $ME < 0$ indicate that the predicted $\theta(h)$ were overestimated and underestimated compared with the measured $\theta(h)$, respectively. Smaller $RMSE$ values indicated better model performance.

All data analysis were also conducted under R [31], and we prepared an R markdown file to reproduce all the analysis in this study ('*Analysis_manuscript.Rmd*' in the GitHub repository: https://github.com/jinshijian/SWRC_MCMC, accessed on 17 June 2022). For more details, please see the "Data and code availability" section below.

3. Results

3.1. Fitted Soil Water Retention Curve by the MH-MCMC Approach and the RETC Program

The results showed that when the complete SWC measurements from saturation to oven dryness were used, both the MH-MCMC approach and the RETC program reasonably simulated the parameters for the VG model, as the fitted curve matched well with the measured data (Figure 3a). Adjusted R^2, RMSE, and ME from the RETC program are similar to that from the MH-MCMC approach (Table 2). Importantly, for soil V (a silt clay loam soil) and soil VIII (a silt loam soil), the RETC program was unable to obtain θ_r values due to extremely small estimates, and 0 values were used for θ_r in the RETC program (Table A2). When using the MH-MCMC approach, however, we were able to obtain θ_r values for soil V and soil VIII (Table A2). MH-MCMC approach showed similar accuracy as the RETC program (i.e., adjusted R^2, RMSE, and ME are similar, see Table 2), but avoiding extreme small value for θ_r, indicating that the MH-MCMC approach developed here is at least as accurate as the RETC program.

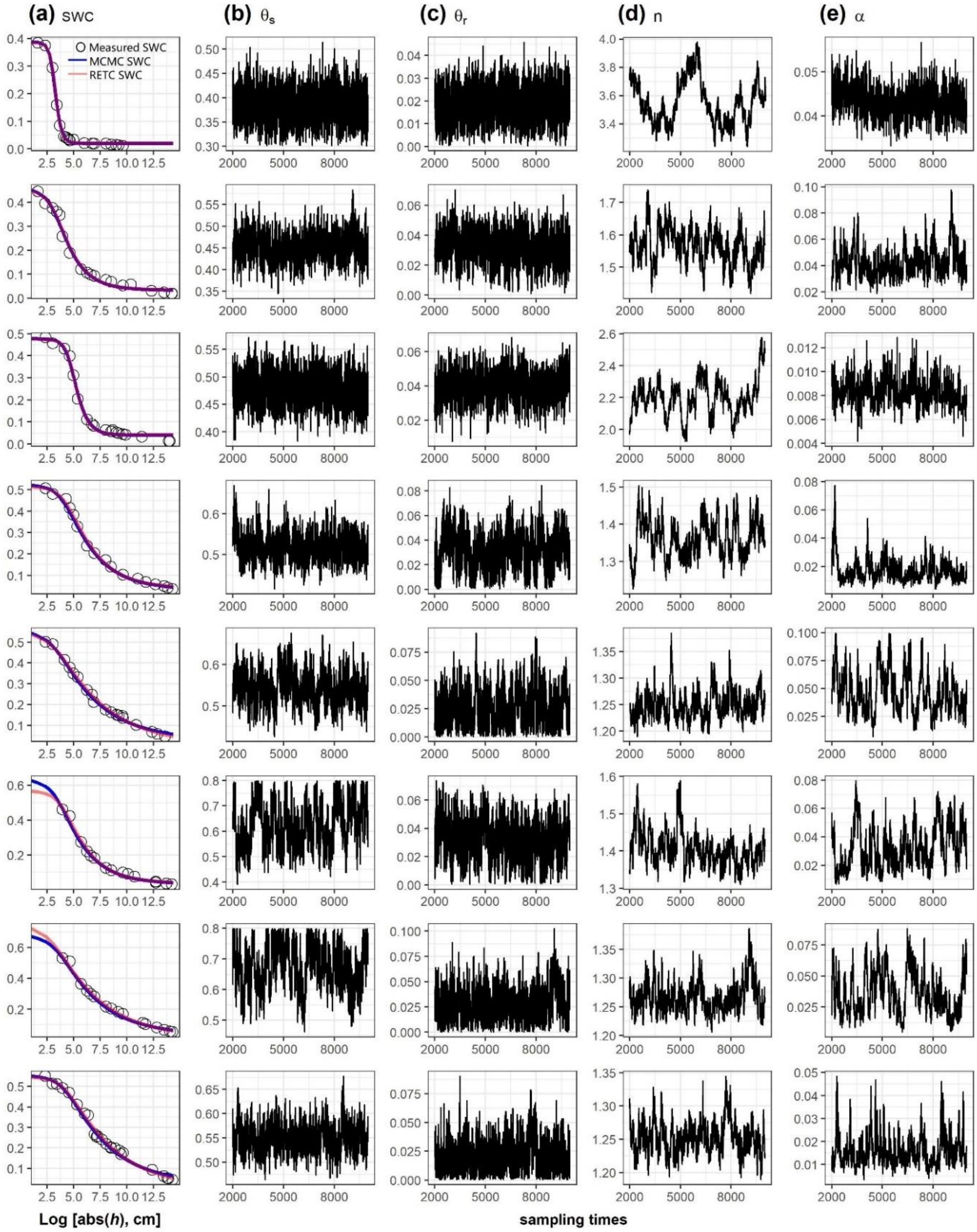

Figure 3. (**a**) Comparison between measured and predicted soil water content using Metropolis–Hastings Markov Chain Monte Carlo (MH-MCMC) approach. (**b**–**e**) Sampling processes of parameters θ_s, θ_r, α, and n, respectively.

The random walk trace plot of θ_S, θ_r, α, and n (Figure 3b–e) indicate convergence in the same direction, albeit with large variations in parameter estimates. A histogram of posteriori estimates of θ_S, θ_r, α, and n (Figure A2) were generally normal distributions. All parameters' auto-correlation decreased as lag-time increased (Figure A3), indicating that the MH-MCMC simulation performs efficiently through the calculations. Our results also showed that with uniform prior distributions of the parameters, the MCMC approach is robust to simulations for all 8 soil types from Lu et al. [24,25] (Figure 3).

3.2. Model Performance if Only 5 Measurements between −60 and −15,000 cm Were Used in the MH-MCMC Approach and the RETC Program

We tested the performance of the RETC program and MH-MCMC approach when only 5 data points (h = −60, −100, −200, −500, and −15,000 cm) were used (Table A1 and red crosses in Figure A4a). Compared with the outputs when all measurements were used to obtain the parameters of the VG model, the accuracy of the RETC program significantly decreased for soil I (a sandy soil), soil II (a sandy loam soil), and soil VII (a silt clay loam soil) when only these 5 measurements were used (Table 2). For example, the adjusted RETC R^2 significantly decreased, while RMSE and ME increased, especially for soil I (sand), the α value was 1.2765 (Table A2), beyond the range (0–1) of the defined α parameter of the VG model (Figures A4–A6, Tables 2 and A2).

Compared with using all SWC measurements, if only 5 data points (h = −60, −100, −200, −500, and −15,000 cm) were used to parameterize the VG model, the MH-MCMC model performance of the soil I (sand) decreased the most, with adjusted R^2 values which decreased from 0.997 to 0.950, RMSE which increased from 0.006 to 0.027, and ME which increased from 0.000 to 0.013 (Table 2). However, even for this worst case scenario, the MH-MCMC approach obtains reliable results for the VG model parameters. Therefore, when fewer measurements were available, the results in this study suggest that MH-MCMC is still a robust way to obtain the parameters of the VG model.

3.3. Model Performance over All 1871 Soils in the NCSS Dataset

As soils from the NCSS dataset only have SWC measured at matric potential between 0 and −15,000 cm, we therefore used the MH-MCMC approach to obtain the VG model parameters based on SWC at metric potential of −60, −100, −330, and −15,000 for all 1871 soils in the NCSS dataset. The random walk trace plot, histogram of posteriori estimates, and the auto-correlation of θ_S, θ_r, α, and n indicate that the MH-MCMC method performs well for all soils. The simulated SWC at h = −60 cm, −100 cm, −330 cm, and −15,000 cm were highly consistent with the measured SWC at corresponding matric potential (Figure 4). The simulated θ_s for the NCSS database matched well with the distribution of measured θ_s (Figure 5a). The distribution of simulated θ_r from the NCSS database was skewed left compared with the SWC at metric potential of −15,000 cm (Figure 5b), indicating that 1) many soils may still retain a non-trivial amount of water at permanent wilting point (h = −15,000 cm), and that 2) it may be pragmatic to extend measured retention values closer to θ_r, or fully dried SWC. The distribution of 1871 simulated n and α values matched well with the predicted n and α values from the UNSODA dataset (Figure 5c,d). All the above comparisons suggested that the MH-MCMC approach developed in this study were able to obtain the VG model parameters for all the 1871 soils in the NCSS dataset.

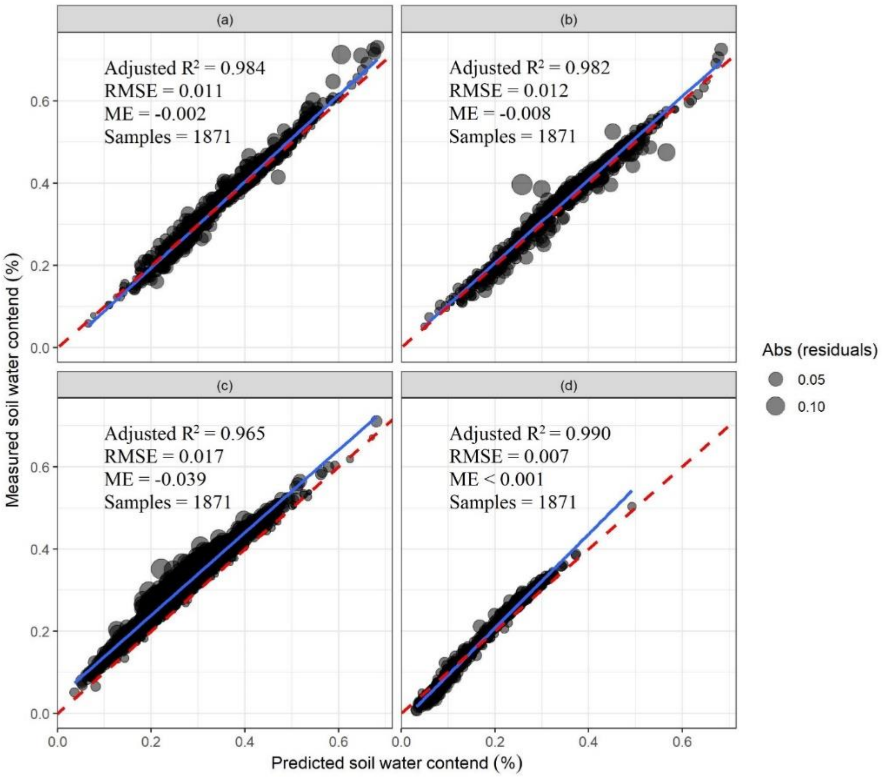

Figure 4. Scatter plot showing the relationship between predicted and measured soil water content at matric potential of −60 cm (**a**), −100 cm (**b**), −330 cm (**c**), and −15,000 cm (**d**), respectively. The comparison is based on all 1871 samples from the National Cooperative Soil Characterization database (NCSS). Note that dot size indicated the absolute values of the residuals between predicted and measured soil water content. The blue lines are regression lines and the dashed red lines are 1:1 lines.

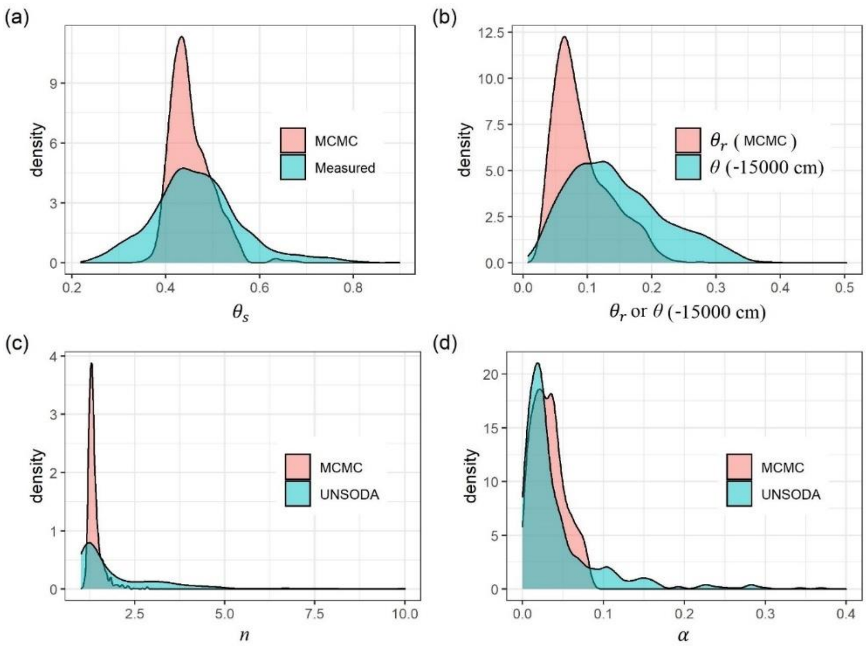

Figure 5. (**a**) Comparison between measured and predicted saturated soil water content (θ_s), (**b**) Comparison between soil water content at matric potential of 15,000 cm and simulated θ_r, (**c**) comparison between simulated n and n from the UNSODA database, and (**d**) between simulated α and α from the UNSODA database.

4. Discussion

Solving the VG model (and its 4 parameters) with limited measurements (sometimes sample size ≤ 5) could introduce parameter uncertainty for various reasons. First, uncertainties may arise from the VG model itself, in the process when describing the SWRC using a relatively simple equation with several parameters which could introduce uncertainties. Second, inherent errors in the collection and measurement of water retention data adds to this compounding uncertainty during model calibration. Finally, standard laboratory water retention methods are generally limited to matric potential between 0 and −15,000 cm [1]. For example, soils from the NCSS dataset (http://ncsslabdatamart.sc.egov.usda.gov/, accessed on 17 June 2022) only measured SWC at matric potential of 0, −60, −100, −330, and −15,000 cm.

Non-convergence and extreme parameters' outputs are common problems when seeking solutions for the VG model parameters using the RETC [12,14,15]. In this study, based on 8 soils from Lu et al.'s dataset [32,33], we find that θ_r cannot be obtained for soil V (a silt loam soil) and soil VIII (a silt clay loam soil) due to extremely small estimates when using the RETC program (Table A2). In this case, the RETC program assumed these low θ_r values to be zero, whereas MH-MCMC derived values were 0.03 and 0.02 for the silt loam and silty clay loam, respectively. However, this discrepancy of 2–3% water content could be hydrologically significant for storage and flow properties, especially when scaled across a soil profile. When only SWC at around matric potential of −60, −100, −200, −500, and −15,000 cm were used, soil I in Lu et al.'s dataset [32,33] in the RETC method produced $\alpha > 1.0$, which is beyond the range (0–1) of the defined α parameter of the VG model. Both issues were well resolved by the MH-MCMC approach developed in this study. The results show that the MH-MCMC approach had superior performance compared to the RETC program in obtaining VG parameters for soils of variable texture.

Measuring the SWRC is usually compromised by time and labor, as such, reliable pedotransfer functions (PTFs) can be useful tools to predict the SWRC based on relatively easy-to-measure soil properties (e.g., soil texture and bulk density). One critical point in improving PTFs is not from new statistical methods but rather from quality data sources [32,33]. In this sense, the VG model parameters from the 1871 soils in the NCSS dataset provide valuable sources for the future development of PTFs to predict the SWRC. Furthermore, the MH-MCMC approach developed in this study is robust in obtaining VG parameters to different soil types and even outperforms traditional least squares-based methods when retention data are limited to $-15{,}000$ cm $< h < -60$ cm. Thus, this new approach could be used to expand the database of VG model parameters in the future. In addition, Jian et al. [34] showed that the widely used ROSETTA model significantly over predicted the near-saturated hydraulic conductivity in urban soils. The MH-MCMC approach developed in this study may be appropriate for characterizing a wide range of soil textures and, as a result, it may help build water retention functions for understudied urban soils.

Even though the newly developed MH-MCMC approach is more robust to solve the VG model when compared to the RETC procedures, it is important to note that the MCMC method has its inherent drawbacks: (1) The sampling points are not independent of each other; the MCMC approach usually takes one sample per N point to alleviate this problem, but it does not solve the problem in and of itself [35]. (2) The mixing time may be very long, i.e., the model may go back and forth in the initial state, within a certain distribution. When the peaks and valleys of the two distributions are too low, the probability of sampling points reaching the peaks and valleys is very low. This can result in the MCMC not being able to sample points that can well characterize the target probability distribution—meaning sampling failure [36]. (3) Even if the sampling is successful, it is difficult to know exactly which moment has reached the stationary distribution [37]. Instead, by periodically adopting the distribution at the current moment, the probability distribution of its sampling is obtained and compared with the target probability distribution to determine how similar the two distributions are, and confirm that a stationary distribution has been reached. For the above reasons, MCMC users should proceed with caution and pay close attention to data quality and parameter posterior distributions before accepting solver solutions.

5. Conclusions

We developed an MH-MCMC code in R for solving VG model parameters, and this MH-MCMC approach performs well across a wide range of soil textures. The results showed that the MH-MCMC approach had good performance when only 5 SWC measured at matric potential of −60cm and −15,000 cm were used to optimize the VG model parameters. The MH-MCMC approach was then used to obtain VG model parameters for 1871 soils in the NCSS dataset. The MH-MCMC code developed in this study provides a useful tool for future usage of solving VG model parameters based on the SWC measurements. The VG model parameters of the 1871 soils in the NCSS dataset likely provide useful information for the future PTF development.

Author Contributions: X.D., C.D. and J.J. conceived and designed the primary analysis; X.D. and C.D. led the drafting of this manuscript. J.R. and Q.W. provided feedback and insights in all phases of model and manuscript development. All authors have read and agreed to the published version of the manuscript.

Funding: Xuan Du was supported by the Soil Moisture Survey and Exploration of Irrigation and Drainage Decision Analysis Tools for Urban Green Space in Yangling project funded by the Yangling Vocational & Technical College, grant number BG2021005. Can Du was supported by the Development and Application of Environmental Friend and Efficient Morel Cultivation Substrate project, grant number BG2021007. Jinshi Jian was supported by the Talent Attracting Supporting Funds by Northwest A & F University under funding number A315022202, and the Strategic Priority Research Program of Chinese Academy of Sciences under contract number, XDA20040202. Jesse Radolinski was funded by the Austrian Academy of Sciences (ÖAW).

Informed Consent Statement: Not applicable.

Data Availability Statement: All data and code to reproduce the results can be found at the following GitHub repository: https://github.com/jinshijian/SWRC_MCMC (accessed on 17 June 2022). Specifically, MCMC_function.R describes functions to simulate and get van Genuchten model parameters (θ_S, θ_r, α, and n) for the soil samples from [24,25] and from the National Cooperative Soil Characterization (NCSS) database using Metropolis–Hastings Markov Chain Monte Carlo (MH-MCMC) approach. Analysis_manuscript.Rmd is a R markdown file including all code to reproduce all analysis and results for this manuscript. Lu_2008_2013_Data.csv includes soil water content under different matric potential gradient from 8 soils [24,25]. NCSS_MCMC2_reorder.csv includes soil water content under matric potential gradient for 1871 soils from the NCSS database, and RosettaFittedVG.csv is the UNSODA van Genuchten model parameters (θ_S, θ_r, α, and n) predicted by the ROSETTA model [26].

Acknowledgments: We appreciate Alexey Shiklomanov for his feedback and suggestions on the coding and manuscript of this project.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Soil texture, volumetric soil water content at selected matric potential of around −60, −100, −200, −500, and −15,000 cm for 8 soils in Lu et al.'s dataset [24,25]. For detailed soil properties and soil water content from saturation to oven dryness conditions, please refer to [24,25] and Lu_2008_2013_Data.csv in the GitHub repository: https://github.com/jinshijian/SWRC_MCMC (accessed on 17 June 2022).

Soil	Texture	Suction Matric (cm)	Volumetric Soil Water Content
I	Sand	−60	0.0446
		−102	0.0304
		−142.9	0.0336
		−403.1	0.0208
		−15,306	0.0112

Table A1. *Cont.*

Soil	Texture	Suction Matric (cm)	Volumetric Soil Water Content
II	Sandy loam	−51	0.2580
		−102	0.1889
		−295.9	0.1213
		−510.2	0.1072
		−15,101.9	0.0536
III	Loam	−61.2	0.4316
		−102	0.3978
		−214.3	0.2041
		−510.2	0.1105
		−15,306	0.0442
IV	Silt loam	−71.4	0.4569
		−102	0.4154
		−204.1	0.3283
		−510.2	0.2365
		−15,306	0.1018
V	Silt clay loam	−61.2	0.4141
		−102	0.3780
		−214.3	0.3315
		−510.2	0.2696
		−15,306	0.1342
VI	Silt loam	−51	0.4615
		−102	0.4229
		−295.9	0.2753
		−510.2	0.2208
		−15,101.9	0.0998
VII	Silt clay loam	−51	0.5306
		−102	0.5095
		−295.9	0.3643
		−510.2	0.3208
		−15,101.9	0.1690
VIII	Silt loam	−50	0.4920
		−95.2	0.4680
		−203.9	0.4104
		−407.9	0.3672
		−15,116.4	0.1764

Table A2. Parameters of van Genuchten model obtained by the Metropolis–Hastings Markov Chain Monte Carlo approach and RETC program.

	MH-MCMC Approach				RETC Program			
Soil	θ_r	θ_S	α	n	θ_r	θ_S	α	n
All measurements were used to obtain VG model parameters								
I	0.0196	0.3870	0.0432	3.7672	0.0206	0.3850	0.0422	3.9701
II	0.0330	0.4532	0.0426	1.5655	0.0339	0.4489	0.0391	1.5755
III	0.0401	0.4770	0.0084	2.2000	0.0418	0.4748	0.0079	2.2908
IV	0.0323	0.5216	0.0170	1.3622	0.0369	0.5141	0.0132	1.3842
V	0.0255	0.5443	0.0461	1.2501	**0.0000** *	0.5378	0.0443	1.2197
VI	0.0333	0.6269	0.0324	1.4044	0.0377	0.5671	0.0185	1.4390
VII	0.0300	0.6719	0.0362	1.2676	0.0119	0.7238	0.0552	1.2370
VIII	0.0213	0.5500	0.0159	1.2521	**0.0000** *	0.5456	0.0144	1.2305
Only 5 measurements at −60, −100, −200, −500, −15,000 cm matric potential were used to obtain VG model parameters								
I	0.0235	0.5527	0.0587	3.7686	0.0101	0.4520	**1.2765**	1.5937
II	0.0569	0.5258	0.0591	1.7282	0.0461	0.8492	0.1995	1.5714
III	0.0475	0.5538	0.0120	2.1971	0.0528	0.4632	0.0071	2.7770
IV	0.0641	0.6305	0.0278	1.4721	0.0734	0.5900	0.0173	1.5191
V	0.0835	0.5568	0.0429	1.3388	0.0481	0.4980	0.0233	1.2816
VI	0.0673	0.5917	0.0237	1.5001	0.0948	0.4855	0.0067	1.9176
VII	0.1047	0.6420	0.0227	1.4061	0.1593	0.5564	0.0061	1.7973
VIII	0.0770	0.6189	0.0431	1.2510	0.0506	0.5308	0.0111	1.2615

* The RETC program was unable to simulate θ_r, therefore 0 value was used in the program.

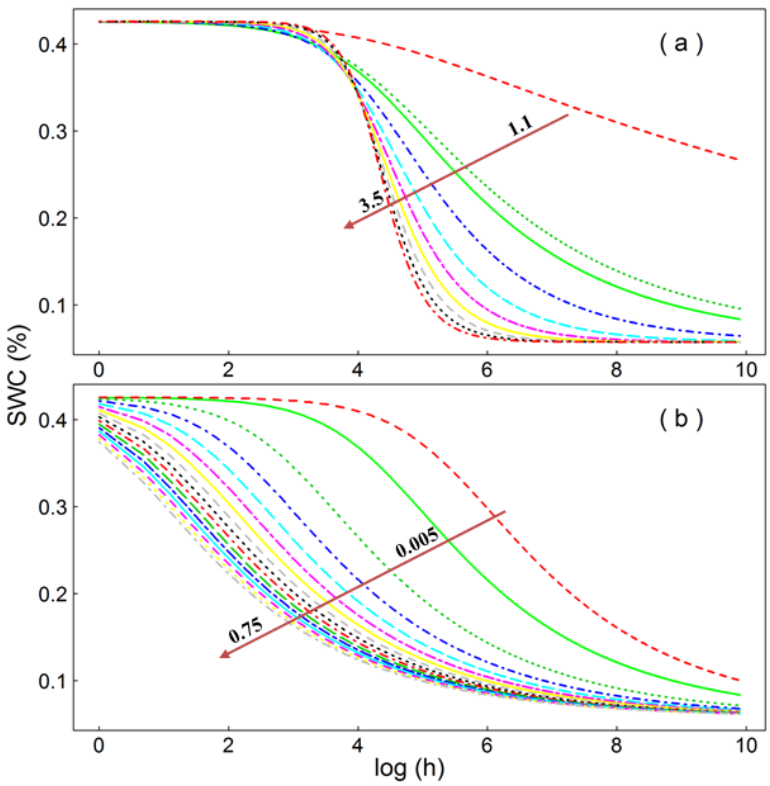

Figure A1. Diagram shows how changing n value affect the soil water retention curve (**a**), and how changing α value affects the soil water retention curve (**b**).

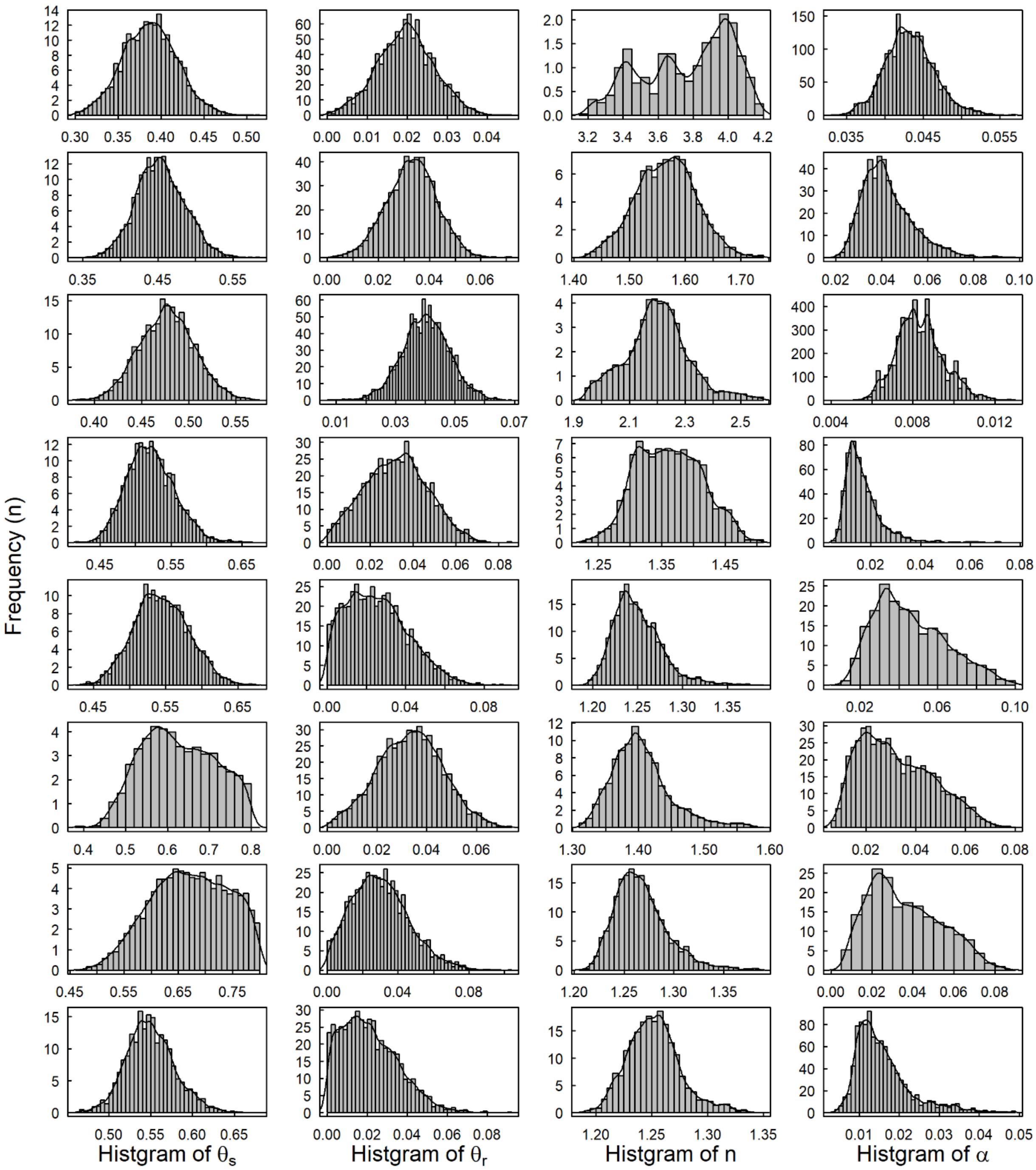

Figure A2. Histogram of posteriori estimates of each parameter (θ_s, θ_r, α, and n) for the van Genuchten model using Metropolis–Hastings Markov Chain Monte Carlo (MH-MCMC) approach.

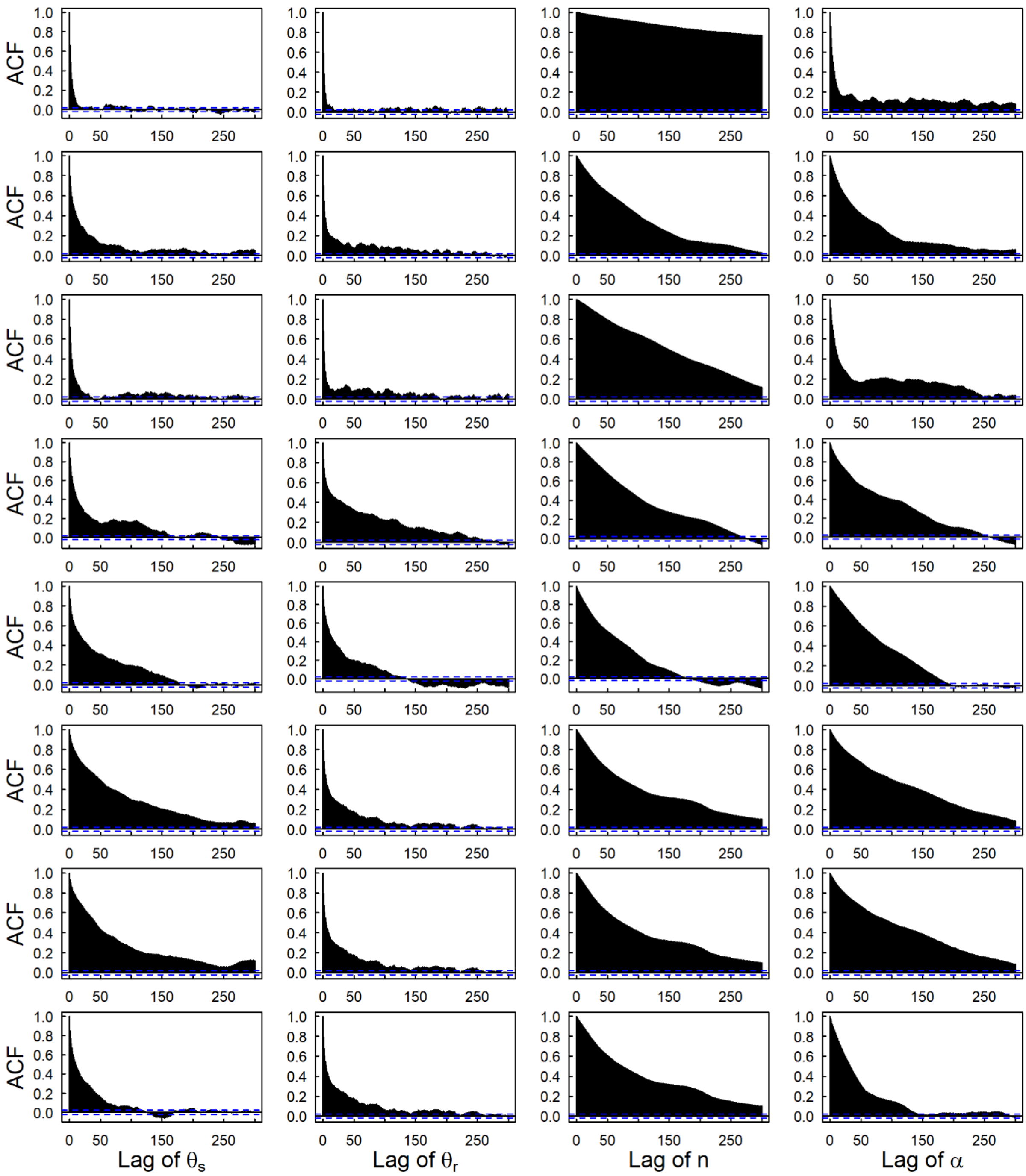

Figure A3. Auto-correlation of each parameter (θ_s, θ_r, α, and n) for the van Genuchten model using Metropolis–Hastings Markov Chain Monte Carlo (MH-MCMC) approach.

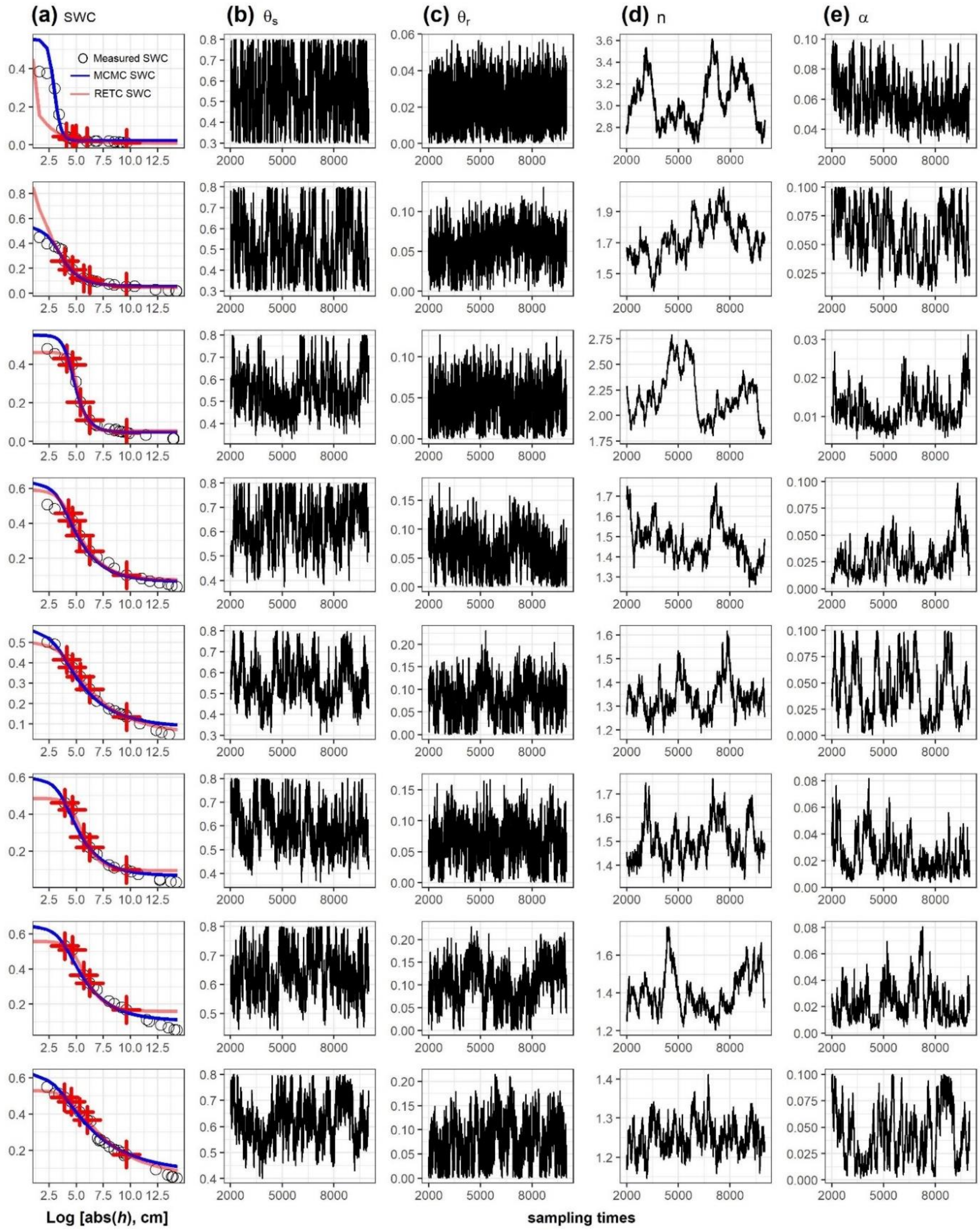

Figure A4. (**a**) Comparison between measured and predicted soil water content using Metropolis–Hastings Markov Chain Monte Carlo (MH-MCMC) approach, (**b**–**e**) Sampling processes of parameters θ_s, θ_r, α, and n, respectively. Note that only soil water content between matric potential gradient of −50 to −16,000 cm (red crosses in (**a**) were used during the MH-MCMC simulation.

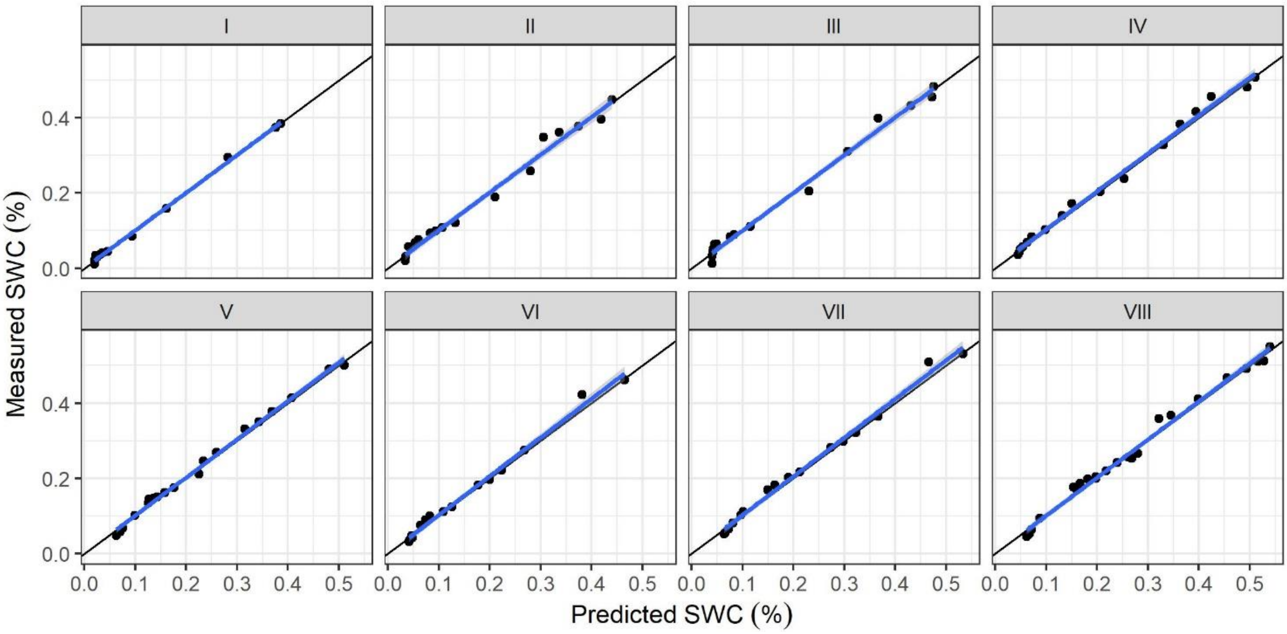

Figure A5. Relationship between measured volumetric soil water content (SWC, $L^3 \cdot L^{-3}$) and predicted SWC ($L^3 \cdot L^{-3}$) for 8 soils from Lu et al. [24,25]. The predicted SWC were from the van Genuchten model parameterized using Metropolis–Hastings Markov Chain Monte Carlo approach. Panels I-VIII corresponding to soil I to soil VIII in Lu et al. [24,25] and Table A1; The black lines are 1:1 lines and the blue lines are regression lines.

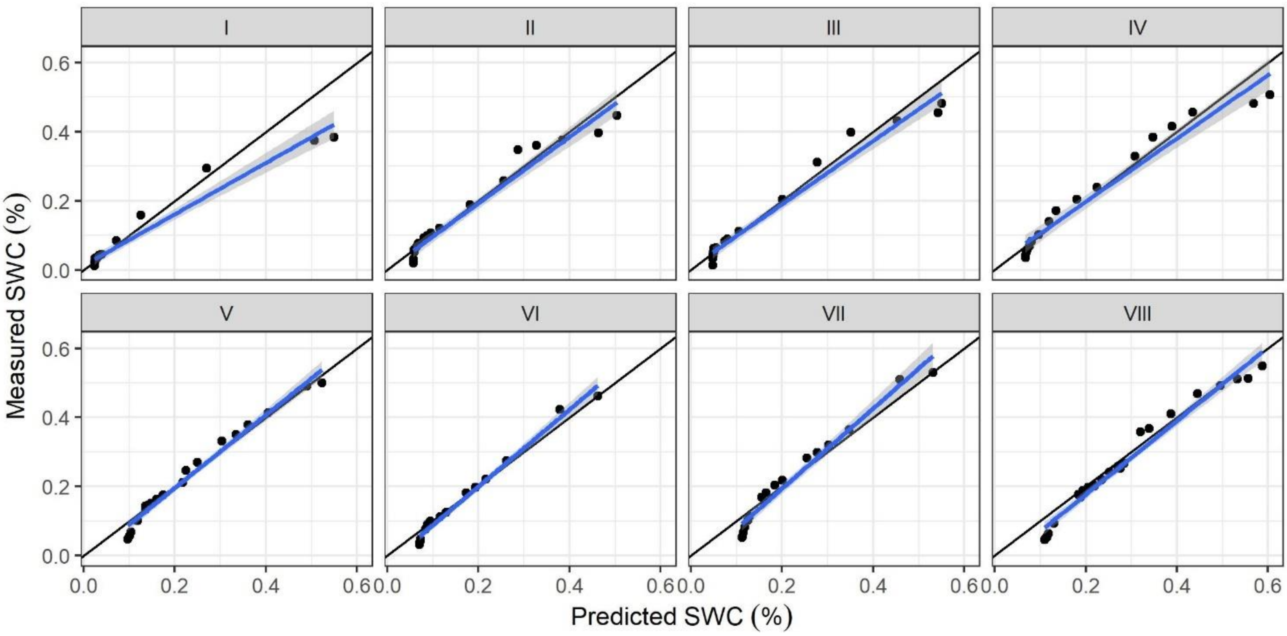

Figure A6. Relationship between measured volumetric soil water content (SWC, $L^3 \cdot L^{-3}$) and predicted volumetric soil water content ($L^3 \cdot L^{-3}$) for 8 soils from Lu et al. [24,25]. The predicted SWC were from the van Genuchten model parameterized using Metropolis–Hastings Markov Chain Monte Carlo approach with only 5 data points (shown as the red crosses in Figure A4). Panels I-VIII corresponding to soil I to soil VIII in Lu et al. [24,25] and Table A1; The black lines are 1:1 lines and the blue lines are regression lines.

References

1. Schelle, H.; Heise, L.; Jänicke, K.; Durner, W. Water Retention Characteristics of Soils over the Whole Moisture Range: A Comparison of Laboratory Methods. *Eur. J. Soil Sci.* **2013**, *64*, 814–821. [CrossRef]
2. Menezes, A.S.; Alencar, T.L.; Júnior, R.N.A.; Toma, R.S.; Romero, R.E.; Costa, M.C.G.; Cooper, M.; Mota, J.C.A. Functionality of the Porous Network of Bt Horizons of Soils with and without Cohesive Character. *Geoderma* **2018**, *313*, 290–297. [CrossRef]
3. Richards, L.A.; Fireman, M. Pressure-Plate Apparatus for Measuring Moisture Sorption and Transmission by Soils. *Soil Sci.* **1943**, *56*, 395–404. [CrossRef]
4. Schindler, U.; Durner, W.; von Unold, G.; Mueller, L.; Wieland, R. The Evaporation Method: Extending the Measurement Range of Soil Hydraulic Properties Using the Air-Entry Pressure of the Ceramic Cup. *J. Plant Nutr. Soil Sci.* **2010**, *173*, 563–572. [CrossRef]
5. Campbell, G.S.; Smith, D.M.; Teare, B.L. Application of a Dew Point Method to Obtain the Soil Water Characteristic. In *Experimental Unsaturated Soil Mechanics*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 71–77.
6. Brooks, R.H.; Corey, A.T. *Hydraulic Properties of Porous Media*; Colorado State University: Fort Collins, CO, USA, 1964; pp. 1–34.
7. Van Genuchten, M.T. A Closed-Form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898. [CrossRef]
8. Fayer, M.J.; Simmons, C.S. Modified Soil Water Retention Functions for All Matric Suctions. *Water Resour. Res.* **1995**, *31*, 1233–1238. [CrossRef]
9. Webb, S.W. A Simple Extension of Two-Phase Characteristic Curves to Include the Dry Region. *Water Resour. Res.* **2000**, *36*, 1425–1430. [CrossRef]
10. Khlosi, M.; Cornelis, W.M.; Douaik, A.; van Genuchten, M.T.; Gabriels, D. Performance Evaluation of Models That Describe the Soil Water Retention Curve between Saturation and Oven Dryness. *Vadose Zone J.* **2008**, *7*, 87–96. [CrossRef]
11. Fredlund, D.G.; Xing, A. Equations for the Soil-Water Characteristic Curve. *Can. Geotech. J.* **1994**, *31*, 521–532. [CrossRef]
12. Groenevelt, P.H.; Grant, C.D. A New Model for the Soil-Water Retention Curve That Solves the Problem of Residual Water Contents. *Eur. J. Soil Sci.* **2004**, *55*, 479–485. [CrossRef]
13. Van Genuchten, M.T.; Leij, F.J.; Yates, S.R. *The RETC Code for Quantifying the Hydraulic Functions of Unsaturated Soils*; U.S. Department of Agriculture, Agricultural Research Service: Riverside, CA, USA, 1991; ISBN 9781848212664.
14. Shi, W.; Li, Q.; Han, Q.; Wang, T.; Chen, X.; Zhang, Y. Performance Analysis of Soil Water Retention Curve Models Based on Different Fitting Optimization Algorithms. *J. Water Res. Water Eng.* **2020**, *31*, 157–165. [CrossRef]
15. Shi, X.; Xu, S.; Liao, K. AM-MCMC Approach to Estimate van Genuchten Model Parameters. *Soils* **2012**, *44*, 345–350.
16. Wang, L.; Huang, C.; Huang, L. Parameter Estimation of the Soil Water Retention Curve Model with Jaya Algorithm. *Comput. Electron. Agric.* **2018**, *151*, 349–353. [CrossRef]
17. Zhang, J.; Wang, Z.; Luo, X. Parameter Estimation for Soil Water Retention Curve Using the Salp Swarm Algorithm. *Water* **2018**, *10*, 815. [CrossRef]
18. Gabrié, M.; Rotskoff, G.M.; Vanden-Eijnden, E. Efficient Bayesian Sampling Using Normalizing Flows to Assist Markov Chain Monte Carlo Methods. *arXiv* **2021**, arXiv:2107.08001.
19. Gao, H.; Zhang, J.; Liu, C.; Man, J.; Chen, C.; Wu, L.; Zeng, L. Efficient Bayesian Inverse Modeling of Water Infiltration in Layered Soils. *Vadose Zone J.* **2019**, *18*, 1–13. [CrossRef]
20. Zhang, J.; Li, W.; Zeng, L.; Wu, L. An Adaptive Gaussian Process-Based Method for Efficient Bayesian Experimental Design in Groundwater Contaminant Source Identification Problems. *Water Resour. Res.* **2016**, *52*, 5971–5984. [CrossRef]
21. Carsel, R.F.; Parrish, R.S.; Jones, R.L.; Hanse, J.L.; Lamb, R.L. Characterizing the Uncertainty of Pesticide Leaching in Agricultural Soils. *J. Contam. Hydrol.* **1988**, *2*, 111–124. [CrossRef]
22. Duan, Q.; Gupta, V. SAC-SMA. *Water Resour. Res.* **1992**, *28*, 1015–1031. [CrossRef]
23. Vrugt, J.A.; Gupta, H.V.; Bouten, W.; Sorooshian, S. A Shuffled Complex Evolution Metropolis Algorithm for Optimization and Uncertainty Assessment of Hydrologic Model Parameters. *Water Resour. Res.* **2003**, *39*, 1–14. [CrossRef]
24. Lu, S.; Ren, T.; Lu, Y.; Meng, P.; Sun, S. Extrapolative Capability of Two Models That Estimating Soil Water Retention Curve between Saturation and Oven Dryness. *PLoS ONE* **2014**, *9*, e113518. [CrossRef]
25. Lu, S.; Ren, T.; Gong, Y.; Horton, R. Evaluation of Three Models That Describe Soil Water Retention Curves from Saturation to Oven Dryness. *Soil Sci. Soc. Am. J.* **2008**, *72*, 1542–1546. [CrossRef]
26. Schaap, M.G.; Leij, F.J.; Van Genuchten, M.T. Rosetta: A Computer Program for Estimating Soil Hydraulic Parameters with Hierarchical Pedotransfer Functions. *J. Hydrol.* **2001**, *251*, 163–176. [CrossRef]
27. Haario, H.; Saksman, E.; Tamminen, J. An Adaptive Metropolis Algorithm. *Bernoulli* **2001**, *7*, 223–242. [CrossRef]
28. Geyer, C.J. Practical Markov Chain Monte Carlo. *Stat. Sci.* **1992**, *7*, 473–483. [CrossRef]
29. Wang, H.; Wang, C.; Wang, Y.; Gao, X.; Yu, C. Bayesian Forecasting and Uncertainty Quantifying of Stream Flows Using Metropolis–Hastings Markov Chain Monte Carlo Algorithm. *J. Hydrol.* **2017**, *549*, 476–483. [CrossRef]
30. Smith, A.F.M.; Roberts, G.O. Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *J. R. Stat. Soc. Ser. B* **1993**, *55*, 3–23. [CrossRef]
31. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.

32. Rotunno Filho, O.C.; de Araujo, A.A.M.; Xavier, L.N.R.; Moreira, D.M.; Di Bello, R.C.; Xavier, A.E.; de Araujo, L.M.N. *Soil Moisture and Soil Water Storage Using Hydrological Modeling and Remote Sensing*; Hartemink, A.E., McBratney, A.B., Eds.; Springer: Cham, Switzerland, 2014; ISBN 9783319060125.
33. Lilly, A.; Lin, H. Using Soil Morphological Attributes and Soil Structure in Pedotransfer Functions. *Dev. soil Sci.* **2004**, *30*, 115–141. [CrossRef]
34. Jian, J.; Shiklomanov, A.; Shuster, W.D.; Stewart, R.D. Predicting Near-Saturated Hydraulic Conductivity in Urban Soils. *J. Hydrol.* **2021**, *595*, 126051. [CrossRef]
35. Martino, L.; Casarin, R.; Leisen, F.; Luengo, D. Adaptive Independent Sticky MCMC Algorithms. *EURASIP J. Adv. Signal Process.* **2018**, *2018*, 5. [CrossRef]
36. Chen, Y.; Dwivedi, R.; Wainwright, M.J.; Yu, B. Fast MCMC Sampling Algorithms on Polytopes. *J. Mach. Learn. Res.* **2018**, *19*, 2146–2231.
37. Hassibi, B.; Hansen, M.; Dimakis, A.G.; Alshamary, H.A.J.; Xu, W. Optimized Markov Chain Monte Carlo for Signal Detection in MIMO Systems: An Analysis of the Stationary Distribution and Mixing Time. *IEEE Trans. Signal Process.* **2014**, *62*, 4436–4450. [CrossRef]