PySWR- A Python Code for Fitting Soil Water Retention Functions

Sama S. Memari[1], T. Prabhakar Clement[1]

[1]Department of Civil, Construction and Environmental Engineering, University of Alabama, Tuscaloosa, Alabama, USA.

Abstract: Soil water retention (SWR) function is an important model that provides an empirical relationship between soil moisture and capillary pressure. We present a simple Python tool for fitting different types of SWR functions to laboratory-measured soil moisture data. Three different optimization methods including the Levenberg-Marquardt (LM) method, Trust Region Reflective (TR) method, and Dog Box (DB) method are considered. We used all three methods to fit the van Genuchten (VG) and Brooks and Corey (BC) models to ten soil moisture datasets. Our results show that the TR method, which allows the user to search for optimal parameter values within a constrained region, is the best approach for fitting these models. We developed a new graphical procedure for evaluating the guesstimates and bounds for different SWR model parameters. Overall, the TR method available in Python, together with the proposed graphical procedure, is an excellent approach for fitting both VG and BC models to soil moisture data.

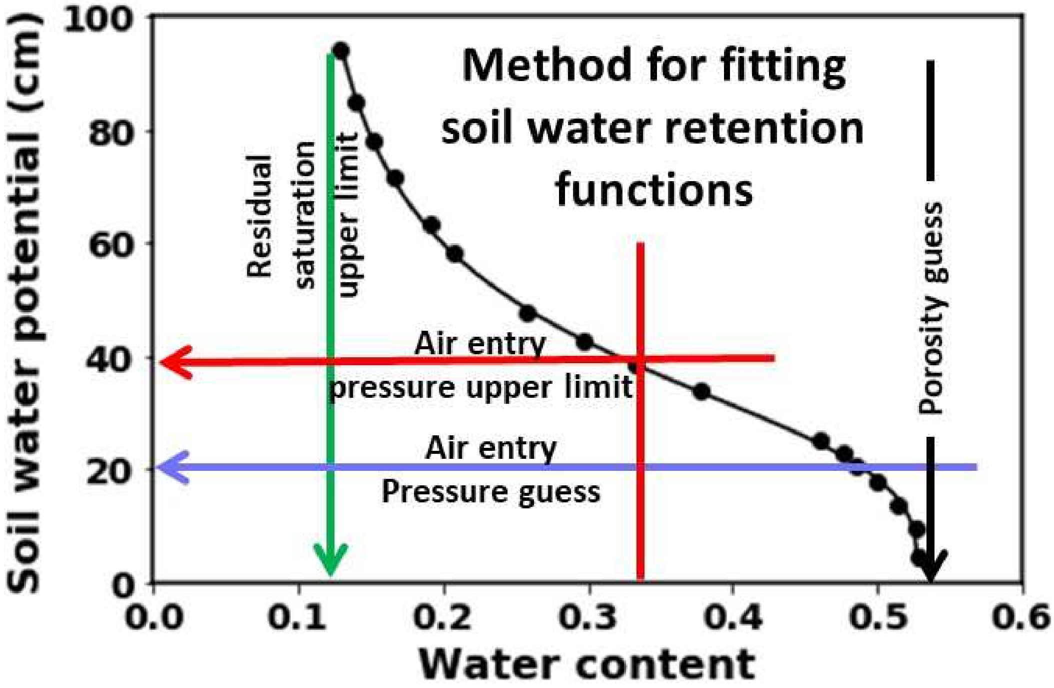
Keywords: soil water retention function; van Genuchten model; Brooks and Corey model; parameter estimation; curve fitting; Python

Author Contribution

SSM developed the code, completed simulations, and co-wrote the manuscript.

TPC developed ideas, debugged the code, and co-wrote the manuscript.

Graphical Abstract



Highlights

- A Python code for fitting soil water retention models to experimental data is provided
- PySWR can fit both van Genuchten and Brooks and Corey water retention models
- PySWR supports two constrained and one unconstrained non-linear fitting methods
- A graphical approach has been developed to provide good initial guesses and parameter bounds
- Trust region reflective method is the best approach for fitting soil water retention models

1. Introduction

A soil water retention (SWR) function is an empirical model that describes the relationship between volumetric water content and soil matric pressure head. This empirical relationship is an important function used in computer simulation tools that are employed for solving practical problems in hydrology and geotechnical engineering fields (Clement et al., 1994; Clement et al., 1996; Tuller et al., 2004; Malaya and Sreedeep, 2012). SWR function characterizes the ability of the soil to store and release water, and is also used for estimating several unsaturated soil properties that are used in hydroclimatic and hydrologic models (Mohanty and Zhu, 2007; Shin et al., 2012). Therefore, both laboratory and field approaches for developing SWR functions have received widespread attention in recent years (Schindler et al., 2012; Masaoka and Kosugi, 2018; Roy et al., 2018; Shokrana and Ghane, 2020).

In the published literature, several analytical models have been suggested for modeling SWR functions and this includes the Brooks and Corey (BC) model (Brooks and Corey, 1964), Fredlund-Xing model (Fredlund and Xing, 1994), Gardner model (Gardner, 1958), Campbell model (Campbell, 1974), and van Genuchten (VG) model (Van Genuchten, 1980), to name a few. Among these models, VG and BC models are the most widely used functions. The parameters in these models are typically identified by fitting these model functions to measured soil moisture data using a nonlinear curve fitting method. Both field and laboratory data have been used in such fitting exercises. For field problems, researchers have employed various types of inverse modeling approaches that utilize unsaturated flow codes, such as HYDRUS, to fit field-observed soil moisture data (Simunek and Van Genuchten, 1999; Wang et al., 2016). Lai and Ren (2016) combined HYDRUS-1D and PEST (a general-purpose parameter estimation software) (Doherty et al., 2010) to determine the effective soil hydraulic parameters at a field site. PEST employs a nonlinear parameter estimation algorithm known as the Gauss-Marquardt-Levenberg method. The results of this study indicated that there are no unique set of average soil properties for fitting water content values measured at a heterogeneous field site. Ket et al. (2018) used a capacitance probe and a dielectric water potential sensor to measure soil water content and water potential, respectively, at a field site. They used HYDRUS-1D to fit the in situ data to indirectly estimate the values of VG parameters for different types of soils. Nascimento et al. (2018) used multiple instruments to measure the values of matric potential and soil moisture levels in a field experiment and then used HYDRUS-1D to estimate the VG model parameters. They concluded that HYDRUS-1D was able to estimate the VG model parameters well, and the values were found to be consistent with laboratory estimates.

For fitting SWR data, researchers have employed different types of nonlinear least square (NLS) algorithms and heuristic search (HS) methods. Several numerical codes have been developed for solving this curve-fitting problem. One of the codes that use an NLS method is the RETC code, and it is used widely for fitting different types of SWR models (Van Genuchten and Yates, 1991). Omuto and Gumbe (2009) used the Gauss-Newton algorithm available in R for fitting soil hydraulic properties used in infiltration and water retention models. Kumar et al. (2020) developed a software tool for fitting BC, VG, and modified-VG models using the Levenberg-Marquardt NLS routine available within the SPSS statistical software package. One of the limitations of using NLS algorithms is that the final solution would depend on the quality of the initial guesses and therefore the estimated model parameters might not be the unique global optimal values. HS algorithms, which are independent of initial conditions, offer a more robust alternative for estimating optimal SWR parameters (Chen et al., 2016). However, HS methods have other numerical parameters that need to be adjusted a priori to obtain valid solutions. This requirement could affect the final output, and also the process of adjusting these parameters can be computationally inefficient (Li et al., 2018; Luo et al., 2018). To avoid the issues related to algorithm-specific parameter adjustments, Zhang et al. (2018) employed a novel salp swarm algorithm (SSA) and used it to fit SWR functions. They also compared the performance of the SSA method with three other methods for fitting SWR functions. Their results indicated that SSA can yield better results. Recently, Guellouz et al. (2020) presented a study where they used the bound optimization by quadratic approximation (BOBYQA) approach to fit a finite difference model, which is based on the Richards' equation, to simulate a field experiment. They analyzed a drainage experiment conducted at field site in Southwestern Tunisia to estimate the VG parameters for the site.

The tools reviewed above require complex computer programs for fitting SWR models, and also all these programs have some computational limitations. The objective of this study is to develop a simple, yet robust, computer tool for fitting VG and BC models to laboratory-measured soil moisture data. The Python module SciPy offers several computationally efficient solvers for fitting a nonlinear function to experimental data. In this study, we developed a Python code, namely PySWR, that employs SciPy for fitting SWR functions. We evaluated the code performance by fitting VG and BC functions to ten experimental datasets available in the literature.

2. Methods

2.1. van Genuchten SWR model

The VG model (Van Genuchten, 1980) is the most widely used SWR model since it is a smooth mathematical function without any discontinuities. This model has been used to describe a broad range of disturbed and undisturbed soils. The model is an explicit analytical function that describes the volumetric water content θ as a function of capillary pressure as:

$$\frac{\boldsymbol{\theta}-\boldsymbol{\theta}_{r}}{\boldsymbol{\theta}_{s}-\boldsymbol{\theta}_{r}}=[\mathbf{1}+(\boldsymbol{\alpha}|\boldsymbol{h}|^{\boldsymbol{n}})]^{-\boldsymbol{m}} \qquad (1)$$

where θ is the volumetric water content (cm^3/cm^3); θ_r is the residual water content (cm^3/cm^3), θ_s is the saturated water content (cm^3/cm^3); h is the capillary pressure head (cm) which is a negative number; α (cm^{-1}) is a parameter that is related to the inverse of the air entry pressure; n is a parameter that is related to the shape of the pore size distribution (Wise, 1992; Wang et al., 2017); and m is typically related to the value of n via the expression: m=1-1/n. The VG model is a two-parameter model and its shape is controlled by the values of α and n. The model parameter α is proportional to the inverse air entry value and its value can range from about 0.005 cm^{-1} for fine clays to about 1 cm^{-1} for coarse sand. The dimensionless value of n controls the shape of the drainage pattern and its value can be as high as 10 for uniform soils (such as well-graded sand) that will have sharp drainage pattern, and it can be as low as 1.1 for heterogeneous soils (such as silty clay) that will have diffused drainage pattern (Wise et al., 1994; Cornelis et al., 2005).

2.2. Brooks and Corey SWR model

Another popular empirical function used for modeling SWR data is the BC model (Brooks and Corey, 1964). This model relates soil moisture value with capillary pressure using the following equations:

$$\boldsymbol{\theta}=\begin{cases}\boldsymbol{\theta}_{r}+(\boldsymbol{\theta}_{s}-\boldsymbol{\theta}_{r})|\boldsymbol{\beta h}|^{-\boldsymbol{\lambda}} & (\boldsymbol{h}<-\mathbf{1}/\boldsymbol{\beta})\\ \boldsymbol{\theta}_{s} & (\boldsymbol{h}\geq-\mathbf{1}/\boldsymbol{\beta})\end{cases} \qquad (2)$$

Where □ (cm^{-1}) is the inverse of air entry value (or bubbling pressure) h_b (cm), λ is a pore size distribution index and other terms are defined above. The BC model is a two-parameter model. Unlike the VG model, the BC model is not a smooth function since it has a discontinuity close to the air entry value, a capillary pressure below which the soil is assumed to be fully saturated. Note, the BC model parameter □ is similar to the VG parameter α. Typically, the value of bubbling pressure h_b (cm) for clay soils is high and can range from about 100 to 200 cm; for sand, it is relatively small and can range from 1 to 10 cm. The pore size distribution index λ is related to the VG parameter n. Lenhard et al. (1989) provided the following analytical expression that approximately relates λ to the value of n:

$$\boldsymbol{\lambda = \frac{m}{1-m}\left(1 - 0.5^{\frac{1}{m}}\right)} \qquad (3)$$

where m=1-1/n. Therefore, similar to n, the parameter λ is also related to the shape of the pore-size distribution. If the pores are relatively uniform the soil will have a sharp drainage pattern (since all the pores will drain at a similar capillary pressure). On the other hand, if the pore size distribution is wide then the soil will have a smooth drainage pattern. The typical value of λ can range from 5 for uniform sand to about 0.1 for highly heterogeneous silty-sandy clay soils (Fuentes et al., 1992; Stankovich and Lockington, 1995).

2.3. Fitting SWR functions to experimental data using non-linear optimization methods

The problem of fitting a SWR model to an experimental dataset can be formulated as a least-squares nonlinear optimization problem, where the model parameters are obtained using a curve-fitting algorithm. Nonlinear curve fitting is a process of minimizing the error between data and model predictions by varying the model parameters over a range of possible values. Here we will employ the following three curved fitting algorithms that are available in the Python SciPy module: Levenberg-Marquardt (LM) algorithm (Levenberg, 1944; Marquardt, 1963), Trust Region Reflective (TR) algorithm (Fletcher, 1980; Sorensen, 1982), and Dogleg algorithm with a rectangular trust region (DB) (Voglis and Lagaris, 2004). In the past, others have used the LM algorithm, which is an unconstrained optimization method, for fitting SWR models (Van Genuchten and Yates, 1991; Zhang et al., 2018). However, the LM method can be inefficient for highly nonlinear problems. For these cases, TR or DB could be a better alternative since they allow the model parameter values to be constrained using a set of user-specified bounds. For example, Le et al. (2017) used a new numerical method to estimate several parameters of a non-linear elastic visco-plastic (EVP) creep model for soft soils. Their numerical approach employed the TR algorithm to fit EVP model parameters. This study also explored some of the limitations of the TR algorithm. As summarized in this study, the TR method approximates the objective function $f(x)$ with a quadratic function $q(s)$ that reflects the behavior of function $f(x)$ in a neighborhood N, which is called the trust-region around a point x_k. The model is "trusted" within a limited region around this current point defined by the trust-region sub-problem. This approach can limit the length of the step as one move from x_k to x_{k+1}. Therefore, the method can be inefficient for very large constrained optimization problems. However, the fitting problem that considered in this study only had two unknown parameters and we did not encounter any computational inefficiencies in all our simulations.

2.4. Experimental data for testing the performance of various curve fitting methods

Ten soil moisture datasets are analyzed in this study. Four of these datasets are taken from Van Genuchten and Yates (1991) study, where these data were used to test the performance of the RETC code for fitting both VG and BC functions. These four RETC soils are labeled as Weld silty clay loam (Jensen and Hanks, 1967), Touchet silt loam (King, 1965), G.E. No. 2 sand (King, 1965), and Sarpy loam (Hanks and Bowers, 1962) (See Table S1 in Supplementary Material for more details about this sample dataset).

Six other datasets were taken from the UNsaturated SOil hydraulic DAtabase (UNSODA). The UNSODA is a public domain resource and it provides a wide range of data for several soils. In this study, we used the UNSODA V2.0 available at this website: https://data.nal.usda.gov/. These soil data are presented in a format that can be directly accessed through Microsoft Access-97 (Schaap et al., 2015). The six datasets selected to study include sandy, silty, loamy, and clayey type soils collected at different field sites (See Table S2 in Supplementary Material for more details about this sample dataset).

3.0. Results and Discussion

The basic source code for the PySWR Python script is available at Github (https://github.com/tpclement/PySWR; see Appendix A for more details). PySWR is a relatively short code that offers three powerful options (LM, TR, and DB algorithms) for fitting both VG and BC models to soil moisture data. The code also supports data visualization and error analysis tools. The experiment data are input to the code in a two-column format (pressure head vs. soil moisture) using a standard EXCEL CSV format (See Table S3 in Supplementary Material for a sample dataset).

3.1 *Van Genucthen model results*

To understand the relative performance of LM, TR, and DB algorithms for fitting the VG model, we first fitted the model to one of the RETC soils (Touchet silt loam (King, 1965)) using all three optimization methods. A standard set of initial guess values for the model parameters, provided by Zhang et al. (2018), was used; these values are summarized in Table 1. The table also provides a generic set of lower and upper bounds given by Zhang et al. (2018); these values were employed when running TR and DB methods. The table also provides a generic set of initial guesses as well as the lower and upper bounds for all BC model parameters.

Table 1. The initial guesses and lower and upper bounds used for various model parameters.

Parameters	**θ_r**	**θ_s**	**VG model**		**BC model**	
			n	**α (cm^{-1})**	**λ**	**□ (cm^{-1})**
Initial guess	0.05	0.4	1	1	0.1	1
Lower bound	0	0	1	0	0	0
Upper bound	1	1	100	100	100	100

The values of VG model parameters estimated by PySWR, literature-derived RETC estimates (Van Genuchten and Yates, 1991), and the computational time taken by all three fitting algorithms are summarized in Table 2. Figure 1 compares experimental data with the fitted model results (note absolute values of soil water potential are plotted in all the figures). The results show that it is almost impossible to distinguish the difference between the curves fitted using the three methods. The data presented in Table 2 also show that all three methods estimated identical parameter values. The code was run on a standard windows-based computer with Intel(R) Core(TM) i5 processor and 8.00 GB memory and all three methods took a fraction of a second to converge.

Table 2. The values of van Genuchten model parameters for Touchet silt loam (King, 1965) estimated using the three fitting methods.

Method	**θ_r**	**θ_s**	**n**	**α (cm^{-1})**	**Comp time (s)**
LM	0.092±(0.004)	0.527±(0.001)	3.5±(0.07)	0.0270±(0.0001)	0.092
TR	0.092±(0.004)	0.527±(0.001)	3.5±(0.07)	0.0270±(0.0001)	0.102
DB	0.092±(0.004)	0.527±(0.001)	3.5±(0.07)	0.0270±(0.0001)	0.130
RETC Code	0.102	0.526	3.5	0.027	---

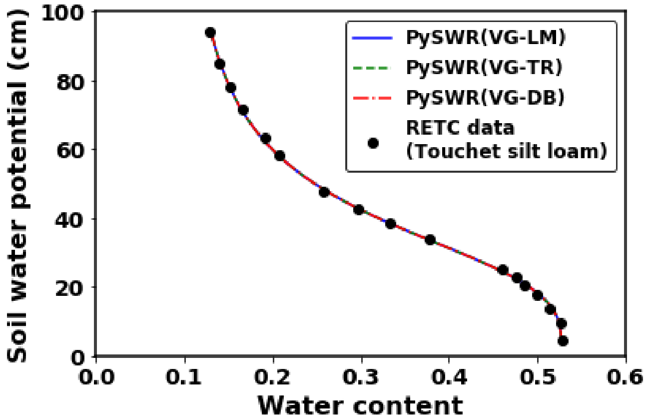


Figure 1. van Genuchten soil water retention function for Touchet silt loam (King, 1965) fitted using the three methods.

The Python tool can also compute the uncertainty (or error) in the estimated values of model parameters (which are the square root of the diagonal entries of the covariance matrix output by the fitting routine). The standard error values for various model parameters estimated by the three fitting methods are summarized in Table 2. Interestingly, the uncertainly estimates computed using all three optimization algorithms are identical.

Since the parameter values estimated by all three fitting methods were identical, for other RETC soils we only report the values estimated using the TR method. We selected the TR method since it is computationally a bit more efficient than the DB method (see Table 2), and it also allowed the user to constrain the parameter space based on our prior knowledge of the parameter values. As illustrated in our later examples, constraining the parameters can have several advantages. Figure 2 presents the model fits for all four RETC soils. In Table 3 the parameter values estimated by PySWR are compared against the values reported in the RETC manual. The figures show that the TR method was able to fit the VG model well for all four RETC datasets. Also, the data shown in the table indicate that the fitted parameter values are close to the values estimated using the RETC code. The sum of square error (SSE) value reported in the table was calculated as the metric to evaluate the difference between the measured and the estimated water content values. The SSE is defined as follows:

$$SSE = \sum_{i=1}^{N} \left(\theta_i^{obs} - \theta_i^{est}\right)^2 \tag{4}$$

Where θ_i^{obs} is the observed data, while θ_i^{est} is the estimated value and N is the total number of measurements in each soil sample. The SSE data show that the TR method provided better fits for most of the soils.

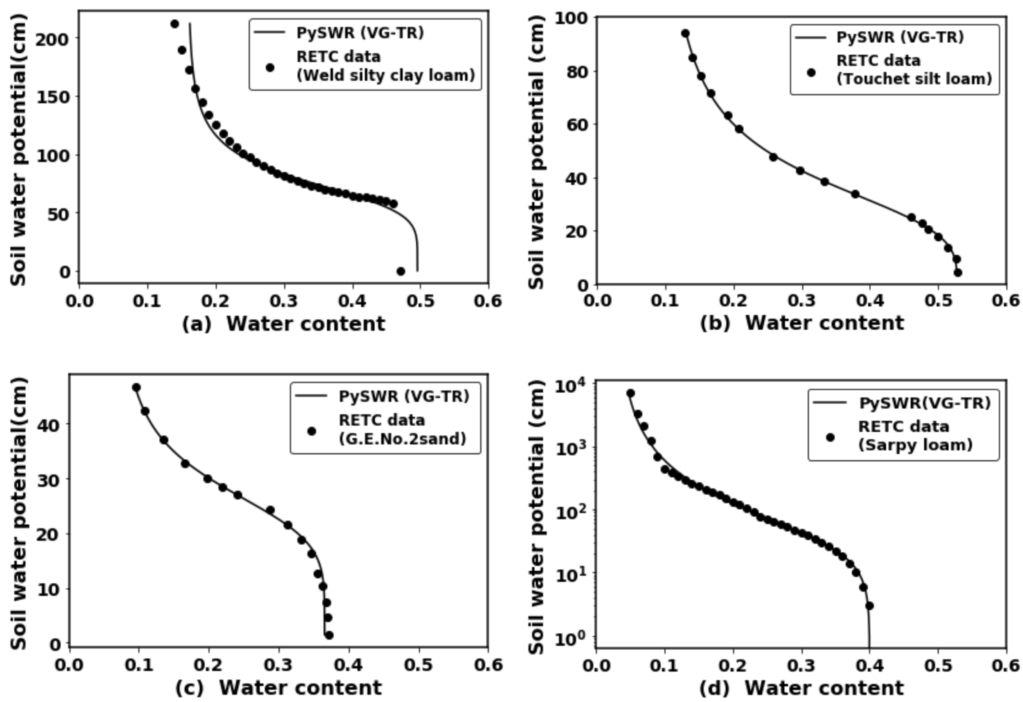


Figure 2. van Genuchten model fits for the four RETC soils fitted using the TR method: (a) Weld silty clay loam (Jensen and Hanks, 1967) (b) Touchet silt loam (King, 1965) (c) G.E.No.2 sand (King, 1965) (d) Sarpy loam (Hanks and Bowers, 1962).

Table 3 The values of van Genuchten model parameters estimated using the TR method for the four RETC soils.

Soil type	**Method**	$\boldsymbol{\theta_r}$	$\boldsymbol{\theta s}$	**n**	$\boldsymbol{\alpha}$ **(cm^{-1})**	**SSE (10^{-3})**
Weld silty clay loam (Jensen and Hanks, 1967)	TR	0.159±(0.006)	0.49±(0.01)	5.4±(0.3)	0.0136±(0.0002)	4.85
	RETC code	0.15	0.49	5.4	0.0136	4.87
Touchet silt loam (King, 1965)	TR	0.092±(0.004)	0.527±(0.001)	3.50±(0.07)	0.0270±(0.0001)	0.10
	RETC code	0.102	0.526	3.59	0.027	0.17
G. E. No.2 sand (King, 1965)	TR	0.069±(0.007)	0.365±(0.002)	5.4±(0.2)	0.0367±(0.0003)	0.23
	RETC code	0.057	0.367	5.0	0.0364	0.34
Sarpy loam (Hanks and Bowers, 1962)	TR	0.031±(0.005)	0.400±(0.002)	1.59±(0.02)	0.027±(0.001)	0.98
	RETC code	0.032	0.400	1.60	0.027	0.99

3.2 Brooks and Corey model results

Similar to the previous section, we used the generic initial guesses and the generic upper and lower bounds provided in Table 1 to fit the BC model to the Touchet silt loam data using all three fitting methods. The model parameter values estimated for Touchet silt loam are summarized in Table 4. The table also provides the optimal values of BC parameters estimated using the RETC code (Van Genuchten and Yates, 1991). These results show that the LM method failed to evaluate good estimates for λ, and even provided an unrealistic negative value for the residual water content. On the other hand, both TR and DB estimated more realistic BC model parameter values. We repeated the fitting exercise for several other soil datasets (details of these soils are discussed in later sections), and for many of these cases, the LM method either failed to converge or estimated unrealistic values. Furthermore, our test simulations indicated that providing better initial guess values and also constraining the parameter values within a narrow range (rather than the broad range provided in Table 1) yielded better results when using the TR and DB methods. Therefore, in the following section, we propose a practical approach for estimating initial guess and upper-and-lower bounds values for various model parameters by graphically analyzing the experimental data.

Table 4. The values of Brooks and Corey model parameters for Touchet silt loam (King, 1965) estimated using the three fitting methods.

Method	$\boldsymbol{\theta_r}$	$\boldsymbol{\theta_s}$	$\boldsymbol{\lambda}$	**□ (cm^{-1})**
LM	-0.67	0.51	0.26	0.05
TR	0	0.510±(0.007)	0.9±(0.2)	0.045±(1e-3)
DB	0	0.510±(0.007)	0.9±(0.2)	0.045±(1e-3)
RETC code	0.018	0.499	1.1	0.037

Figure 3 summarizes the details of the proposed graphical approach for evaluating better initial guesses and parameter bounds. We present the data analysis steps for the Touchet silt loam (King, 1965) dataset to demonstrate this intuitive graphical approach. As a first step, we estimated the initial guess value for porosity by drawing a vertical line connecting a few data points which are close to maximum water content. As shown in Figure 3, this line (black line) intersected the x-axis at the moisture content value of about 0.52, which will be our initial guess for the value of saturated water content (or porosity). We then perturbed this porosity value by about 25% on either side to estimate the lower and upper bounds for porosity as 0.40 to 0.65, respectively.

To estimate the initial guess value of the air entry pressure, we evaluated a transition point where the soil started to drain sharply (i.e., the water content started to decrease sharply from the maximum saturation level) and a horizontal line (blue-line) was drawn through this point and the line intersected the y-axis at the capillary pressure value of about 20 cm, which was assumed to be the guess value of the air entry pressure. To estimate the upper bound for the air entry pressure, we identified an inflection point that normally occurs somewhere between 25% to 50%

of the drainable porosity (this range is an estimate based on our experience of analyzing multiple SWR datasets including the ten datasets presented in this study). For the loam soil, the inflection point is located close to the water content value of 0.35 (as marked by the vertical red line without an arrow). We then drew a horizontal line (red line) going through the inflection point and estimated the upper bound for the air entry pressure as 40 cm for the loam soil (See Figure 3). The lower bound for the air entry pressure was always assumed to be 1 cm (which is an extremely low value, typically observed for coarse sands).

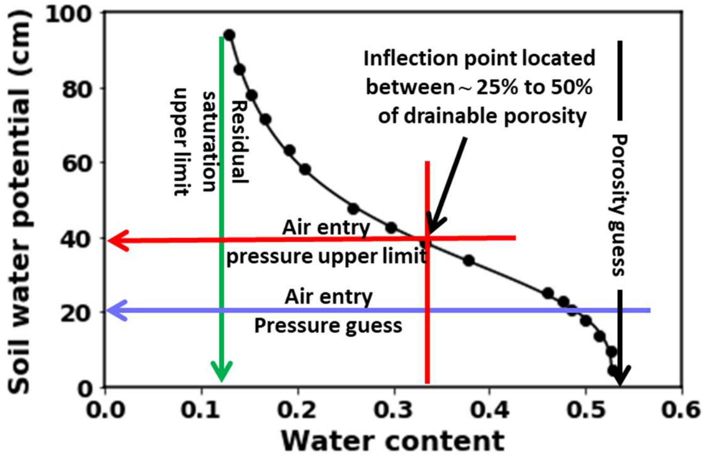


Figure 3. Graphic approach for estimating the initial guess values, and lower and upper bounds for VG and BC model parameters.

To estimate the value of the upper bound of the residual saturation, a vertical line (green line) was drawn that connected a few data points that have very low water content values. This line intersected the x-axis at the water content value of 0.12, and this value was assumed as the upper bound for residual saturation. Since the value of residual water content will always be tending towards zero, we assumed zero as the lower bound for all soils. However, if a better value of residual saturation is available, the user can always use that value as the lower bound. The best initial guess for residual saturation was then estimated as the midpoint between the upper and lower bound values; for the loam soil this value is estimated to be 0.06.

The typical value of λ would range from 5 for uniform material (such as uniform sand), to a low value of about 0.1 for highly heterogeneous silty-clay materials. The initial value for λ was always assumed to be 1, which is close to the logarithmic midpoint of the range of possible λ values. We analyzed all four RETC soils using the proposed graphical approach and estimated initial guesses and upper and lower bounds, and the data are summarized in Table 5.

Table 5. Initial guess values and lower and upper bounds for Brooks and Corey parameters estimated using the proposed graphical approach for the four RETC soils (the values are organized as θ_r, θ_s, λ, □ (cm^{-1}))

Soil type	Initial guess	Lower bound	Upper bound
Weld silty clay loam (Jensen and Hanks, 1967)	0.07, 0.47, 1.0, 0.016	0, 0.35, 0.1, 1	0.14, 0.58, 5, 0.012
Toucher silt loam (King 1965)	0.06, 0.52, 1.0, 0.055	0, 0.39, 0.1, 1	0.12, 0.65, 5, 0.025
G. E. No.2 sand (King, 1965)	0.045, 0.37, 1.0, 0.083	0, 0.28, 0.1, 1	0.09, 0.46, 5, 0.04
Sarpy loam (Hanks and Bowers, 1962)	0.035, 0.40, 1.0, 0.1	0, 0.3, 0.1, 1	0.07, 0.5, 5, 0.0025

We employed the values given in Table 5 to fit the BC model to all four RETC soil datasets and the results are shown in Figure 4.

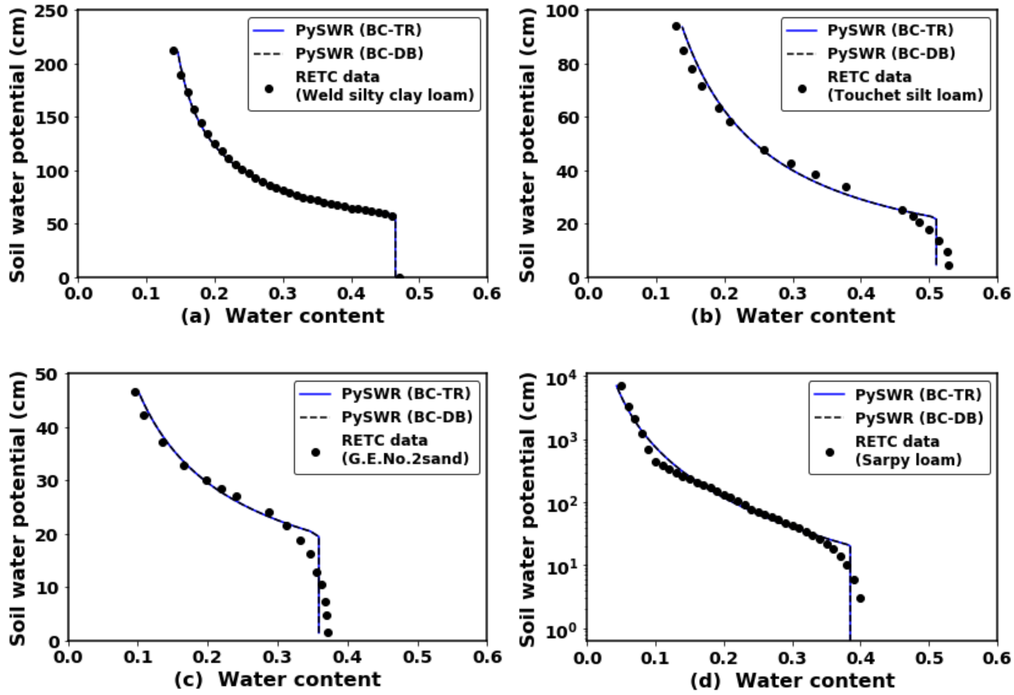


Figure 4. Brooks and Corey model fit for the four RETC soils fitted using the TR and DB methods: (a) Weld silty clay loam (Jensen and Hanks, 1967) (b) Touchet silt loam (King, 1965) (c) G.E.No.2 sand (King, 1965) (d) Sarpy loam (Hanks and Bowers, 1962).

The results show that both TR and DB methods fitted the data well. In Table 6 we only provide the TR results and compare them with the values estimated by the RETC code. These data show that the model parameter estimated by PySWR are close to the RETC estimates. Also, the estimated values of SSE in Table 6 indicate that the TR method performed similar or better when compared to RETC code results, for all four datasets.

Table 6. The values of Brooks and Corey model parameters estimated by TR method for the four RETC soils.

Soil type	Method	θ_r	θ_s	λ	□ (cm^{-1})	SSE (10^{-3})
Weld silty clay loam (Jensen and Hanks, 1967)	TR	0.112±(0.003)	0.470±(0.003)	1.83±(0.03)	0.01739±(6e-5)	0.21
	RETC code	0.11	0.46	1.89	0.017	0.21
Toucher silt loam (King 1965)	TR	0.00	0.510±(0.007)	0.9±(0.2)	0.045±(1e-3)	2.25
	RETC code	0.018	0.499	1.1	0.037	3.67
G. E. No.2 sand (King, 1965)	TR	0.00	0.358±(0.004)	1.5±(0.5)	0.049±(1e-3)	1.56
	RETC code	0.00	0.352	1.7	0.046	3.54
Sarpy loam (Hanks and Bowers, 1962)	TR	0.00	0.380±(0.004)	0.38±(0.04)	0.044±(2e-3)	5.38
	RETC code	0.00	0.380	0.38	0.044	5.39

3.3 Comparison of the efficiency of different non-linear fitting approaches

To understand the relative efficiency of the three fitting approaches, we artificially perturbed the Touchet silt loam (King, 1965) data (the perturbed dataset is given in Table S4, see Supplementary Material) by introducing some random noise to the data. We employed the graphical approach for reevaluating the initial guesses and parameter bounds for the noisy dataset and the results are summarized in Table S5 (see Supplementary Material). We then used all three methods to fit both VG and BC models to this noisy dataset. The estimated model parameter values are summarized in Table 7; note, in the table we only report the values estimated by TR because the LM method failed, and TR and DB methods generated similar results. Figure 5 shows the model profiles fitted using the TR and DB methods. The figure clearly shows that both TR and DB fits were almost identical. The most interesting result of this efficiency test was that the LM method not only failed to fit the BC model (which should be expected) but also failed to fit the VG model when the data was noisy.

Our simulation results also indicated that for most cases the TR method is a bit more computationally efficient than the DB approach (e.g., see Table 2). We completed additional sensitivity simulations by perturbing the initial guess values; the results indicated that for some soils the DB method can be relatively more sensitive to initial guess values when compared to the TR method. Overall, we found the TR method as the most robust approach for fitting both VG and BC models. Therefore, in the following validation section, we only present the results for the TR method.

Table 7. The values of van Genuchten and Brooks and Corey model parameters estimated for the noisy Touchet silt loam (King 1965) data (note, we only report TR results since LM failed to converge and DB estimates were identical to TR).

Method	**θ_r**	**θ_s**	**n or λ**	**α or □ (cm^{-1})**
VG-TR	0	0.51±(0.05)	3±(2)	0.025±(0.005)
BC-TR	0	0.50±(0.04)	0.9±(1)	0.045±(0.008)

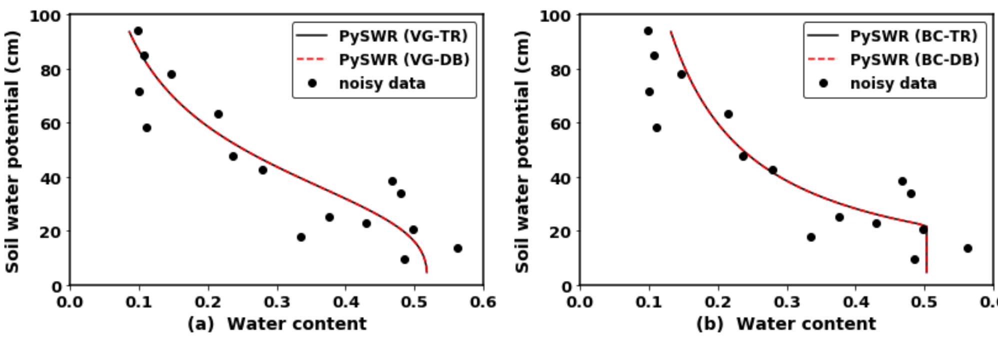


Figure 5. TR and DB fits for the noisy Touchet silt loam data (King, 1965): (a) van Genuchten function model results and (b) Brooks and Corey model results.

3.4 Validation of the code performance using additional datasets

To further test the performance of the PySWR code, we used the code to fit both VG and BC models to six different UNSODA datasets. We first analyzed these experimental data using the proposed graphical approach and evaluated the initial guesses and bounding values for all the model parameters. These values are summarized in Table S6 (see Supplementary Material).

We used the TR method to fit the VG model to the six UNSODA soils and the fitted model profiles are compared with the experimental data in Figure 6. The figures show that the PySWR code was able to fit all UNSODA datasets well. The estimated model parameter values are compared against the SSA (Zhang et al., 2018) and RETC code results in Table S7 (see Supplementary Material). From the values of SSE, summarized in Table S7 it can be observed that the TR method was able to provide better fits with relatively low SSE values when compared to RETC fits.

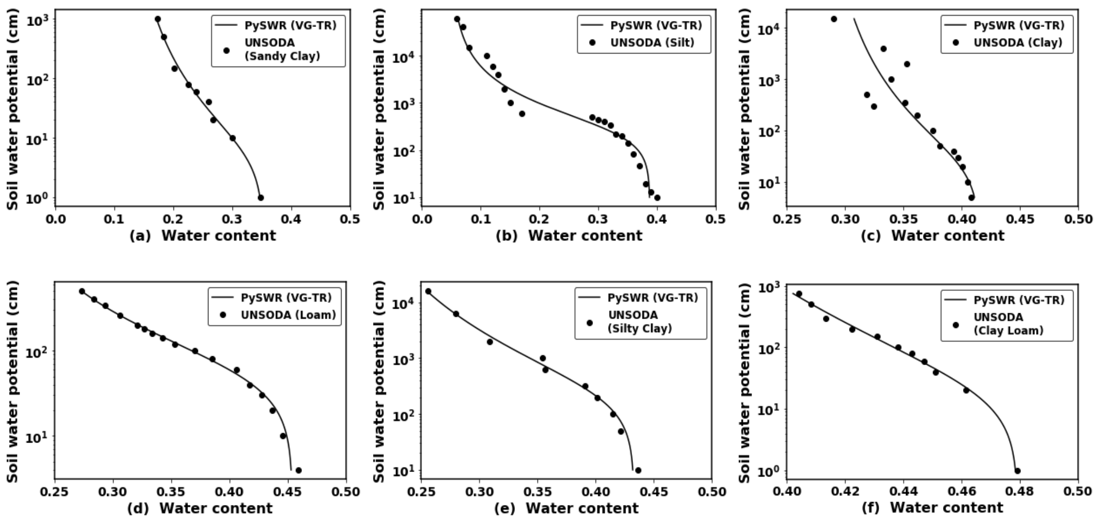


Figure 6. van Genuchten model fits for the six UNSODA soils fitted using the TR method: (a) Sample 1102 (Sandy Clay), (b) Sample 1330 (Silt), (c) Sample 1162 (Clay), (d) Sample 2400 (Loam), (e) Sample 1361 (Silty Clay), (f) Sample 1173 (Clay Loam).

The TR method was then used to fit the BC model to all UNSODA soils. The model profiles fitted by the PySWR code are compared with experimental data in Figure 7. The figure shows that the TR method was able to fit the BC model to all six datasets. Furthermore, the fitted model parameter values are summarized in Table S8 (see Supplementary Material). The values presented in the table (see S8) show that the code was able to estimate realistic model parameters.

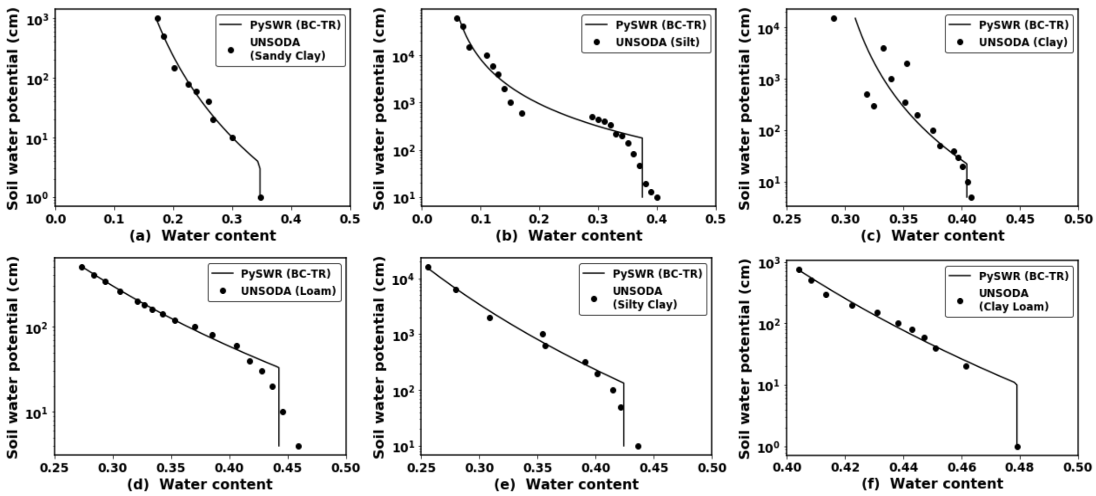


Figure 7. Brooks and Corey model fit for six UNSODA soils fitted using the TR method: (a) Sample 1102 (Sandy Clay), (b) Sample 1330 (Silt), (c) Sample 1162 (Clay), (d) Sample 2400 (Loam), (e) Sample 1361 (Silty Clay), and (f) Sample 1173 (Clay Loam).

As discussed in the aforementioned sections, unlike the VG model, the BC model is not a smooth function and has a discontinuity close to the air-entry value. Comparisons of experimental data shown in Figure 6 and Figure 7 indicate that the initial drainage pattern was fairly smooth for all six UNSODA soils. As expected, the sharp transition region near the air entry value resulted in the BC function not fitting some of the data points located near high water content values. The VG model, which simulated a smoother drainage pattern provided better fits for all six UNSODA datasets. Our test simulations indicated that for most of our cases, the VG model performed well even when generic initial guesses and generic upper and lower bounds were used. On the other hand, the BC model required better initial values and narrower bounds to obtain meaningful results. Overall, the VG model was a better function for describing the UNSODA data.

4. Conclusions

We present the details of a Python code, PySWR, for fitting VG and BC models to soil moisture data. PySWR provides options to use several non-linear least-squares fitting methods, including LM, TR, and DB methods, available in the Python SciPy module. The results show that all three methods were able to fit the VG model to the four RETC soil datasets. However, further analysis indicated that the LM method failed to fit the VG model when some random noise was introduced into the data. The LM method also failed to fit the BC model to all the experimental datasets considered in this study. The TR and DB methods were found to be much better alternatives since they allowed the user to constrain the bounds of various model parameters, thus limiting the search within a feasible range. The efficiency of these methods can be improved by providing good initial guesses, and better upper and lower parameter bounds. The graphical method proposed in this study is an intuitive practical approach for evaluating good guesstimates and parameter bounds. The performance of the DB method was always comparable to the TR method; however, we recommend the use of the TR method since it was relatively less sensitive to variations in initial guess values, and it was also a bit more computationally efficient than the DB method. Our results show that PySWR is an excellent tool for analyzing SWR data. The PySWR code has tools for estimating parameter error, and it also supports various plotting routines for comparing model-fitted SWR curves with experimental data. Overall, PySWR is a useful tool for fitting both VG and BC models to experimental data.

Acknowledgment

This work was, in part, supported by NOAA/UCAR under contract NA18NWS4620043B, and NSF Award Number: 2019561. Both SSM and TPC co-wrote the manuscript and contributed equally to this effort. We like to thank Computers & Geosciences journal reviewers for their helpful review comments.

Computer Code Availability

The PySWR code was jointly developed by the authors and their contact details are given above. The Python code was developed using the Spider interface and was tested on a Windows computer with Intel (R) i5 processor and 8.00 GB memory. The code is available at https://github.com/tpclement/PySWR.

References

Brooks, R.H. and Corey, A.T. 1964. Hydraulic properties of porous media and their relation to drainage design. Transactions of the ASAE 7(1), 26-0028.

Campbell, G.S. 1974. A simple method for determining unsaturated conductivity from moisture retention data. Soil science 117(6), 311-314.

Chen, G., Jiao, L. and Li, X. 2016. Sensitivity analysis and identification of parameters to the Van Genuchten equation. Journal of Chemistry 2016, 8.

Clement, T., Hooker, B. and Skeen, R. 1996. Macroscopic models for predicting changes in saturated porous media properties caused by microbial growth. Ground Water 34(5), 934-942.

Clement, T., Wise, W.R. and Molz, F.J. 1994. A physically based, two-dimensional, finite-difference algorithm for modeling variably saturated flow. Journal of Hydrology 161(1-4), 71-90.

Cornelis, W.M., Khlosi, M., Hartmann, R., Van Meirvenne, M. and De Vos, B. 2005. Comparison of unimodal analytical expressions for the soil-water retention curve. Soil Science Society of America Journal 69(6), 1902-1911.

Doherty, J.E., Hunt, R.J. and Tonkin, M.J. 2010. Approaches to highly parameterized inversion: A guide to using PEST for model-parameter and predictive-uncertainty analysis. US Geological Survey Scientific Investigations Report 5211, 71.

Fletcher, R. (1980) Practical Methods Of Optimization: Unconstrained Optimization, John Wiley & Sons.

Fredlund, D.G. and Xing, A. 1994. Equations for the soil-water characteristic curve. Canadian geotechnical journal 31(4), 521-532.

Fuentes, C., Haverkamp, R. and Parlange, J.-Y. 1992. Parameter constraints on closed-form soilwater relationships. Journal of hydrology 134(1-4), 117-142.

Gardner, W.R. 1958. Some steady state solutions of unsaturated moisture flow equations with application to evaporation from a water table. Soil Science 85(4), 228-232.

Guellouz, L., Askri, B., Jaffré, J. and Bouhlila, R. 2020. Estimation of the soil hydraulic properties from field data by solving an inverse problem. Scientific Reports 10(1), 1-11.

Hanks, R. and Bowers, S. 1962. Numerical solution of the moisture flow equation for infiltration into layered soils. Soil Science Society of America Journal 26(6), 530-534.

Jensen, M.E. and Hanks, R.J. 1967. Nonsteady-state drainage from porous media. Proceedings of the American Society of Civil Engineers, Journal of the Irrigation and Drainage Division 93(IR3), 209-231.

Ket, P., Oeurng, C. and Degré, A. 2018. Estimating Soil Water Retention Curve by Inverse Modelling from Combination of In Situ Dynamic Soil Water Content and Soil Potential Data. Soil Systems 2(4), 55.

King, L. 1965. Description of soil characteristics for partially saturated flow. Soil Science Society of America Journal 29(4), 359-362.

Kumar, N., Poddar, A. and Shankar, V. 2020. Nonlinear Regression for Identifying the Optimal Soil Hydraulic Model Parameters. 979, 25-34.

Lai, J. and Ren, L. 2016. Estimation of effective hydraulic parameters in heterogeneous soils at field scale. Geoderma 264, 28-41.

Le, T.M., Fatahi, B., Khabbaz, H. and Sun, W. 2017. Numerical optimization applying trust-region reflective least squares algorithm with constraints to optimize the non-linear creep parameters of soft soil. Applied Mathematical Modelling 41, 236-256.

Lenhard, B.R.J., Parker, J.C. and Mishra, S. 1989. On the correspondence between Brooks Corey and van Genuchten models. Journal of Irrigation and Drainage Engineering 115(4), 744-751.

Levenberg, K. 1944. A method for the solution of certain non-linear problems in least squares. Quarterly of applied mathematics 2(2), 164-168.

Li, Y.B., Liu, Y., Nie, W.B. and Ma, X.Y. 2018. Inverse modeling of soil hydraulic parameters based on a hybrid of vector-evaluated genetic algorithm and particle swarm optimization. Water 10(1), 84.

Luo, X., Cao, L., Wang, L., Zhao, Z. and Huang, C. 2018. Parameter identification of the photovoltaic cell model with a hybrid Jaya-NM algorithm. Optik 171, 200-203.

Malaya, C. and Sreedeep, S. 2012. Critical review on the parameters influencing soil-water characteristic curve. Journal of Irrigation and Drainage Engineering 138(1), 55-62.

Marquardt, D.W. 1963. An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics 11(2), 431-441.

Masaoka, N. and Kosugi, K.i. 2018. Improved evaporation method for the measurement of the hydraulic conductivity of unsaturated soil in the wet range. Journal of Hydrology 563, 242-250.

Mohanty, B.P. and Zhu, J. 2007. Effective hydraulic parameters in horizontally and vertically heterogeneous soils for steady-state land–atmosphere interaction. Journal of Hydrometeorology 8(4), 715-729.

Nascimento, Í.V.d., Assis Júnior, R.N.d., Araújo, J.C.d., Alencar, T.L.d., Freire, A.G., Lobato, M.G.R., Silva, C.P.d., Mota, J.C.A. and Nascimento, C.D.V.d. 2018. Estimation of van Genuchten Equation Parameters in Laboratory and through Inverse Modeling with Hydrus-1D. Journal of Agricultural Science 10(3), 102.

Omuto, C. and Gumbe, L. 2009. Estimating water infiltration and retention characteristics using a computer program in R. Computers & Geosciences 35(3), 579-585.

Roy, D., Jia, X., Steele, D.D. and Lin, D. 2018. Development and Comparison of Soil Water Release Curves for Three Soils in the Red River Valley. Soil Science Society of America Journal 82(3), 568-577.

Schaap, A., Leij, M., Wösten, F.J. and M, J.H. 2015 UNSODA 2.0: Unsaturated Soil Hydraulic Database. Database and program for indirect methods of estimating unsaturated hydraulic properties.

Schindler, U., Mueller, L., da Veiga, M., Zhang, Y., Schlindwein, S. and Hu, C. 2012. Comparison of water-retention functions obtained from the extended evaporation method and the standard methods sand/kaolin boxes and pressure plate extractor. Journal of Plant Nutrition and Soil Science 175(4), 527-534.

Shin, Y., Mohanty, B.P. and Ines, A.V. 2012. Soil hydraulic properties in one-dimensional layered soil profile using layer-specific soil moisture assimilation scheme. Water Resources Research 48(6).

Shokrana, M.S.B. and Ghane, E. 2020. Measurement of soil water characteristic curve using HYPROP2. MethodsX 7, 100840.

Simunek, J. and Van Genuchten, M.T. 1999. Using the HYDRUS-1D and HYDRUS-2D codes for estimating unsaturated soil hydraulic and solute transport parameters. Characterization and measurement of the hydraulic properties of unsaturated porous media 1, 523-521.

Sorensen, D.C. 1982. Newton's method with a model trust region modification. SIAM Journal on Numerical Analysis 19(2), 409-426.

Stankovich, J. and Lockington, D. 1995. Brooks-Corey and van Genuchten soil-water-retention models. Journal of Irrigation and drainage Engineering 121(1), 1-7.

Tuller, M., Or, D. and Hillel, D. 2004. Retention of water in soil and the soil water characteristic curve. Encyclopedia of Soils in the Environment 4, 278-289.

Van Genuchten, M.T. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils Soil science society of America journal 44(5), 892-898.

Van Genuchten, M.T. and Yates, F.J.L.a.S.R. 1991 The RETC Code for Quantifying the Hydraulic Functions of Unsaturated Soils.

Voglis, C. and Lagaris, I. 2004 A rectangular trust region dogleg approach for unconstrained and bound constrained nonlinear optimization.

Wang, J.-P., Hu, N., François, B. and Lambert, P. 2017. Estimating water retention curves and strength properties of unsaturated sandy soils from basic soil gradation parameters. Water Resources Research 53(7), 6069-6088.

Wang, T., Franz, T.E., Yue, W., Szilagyi, J., Zlotnik, V.A., You, J., Chen, X., Shulski, M.D. and Young, A. 2016. Feasibility analysis of using inverse modeling for estimating natural groundwater recharge from a large-scale soil moisture monitoring network. Journal of Hydrology 533, 250-265.

Wise, W.R. 1992. A new insight on pore structure and permeability. Water Resources Research 28(1), 189-198.

Wise, W.R., Clement, T. and Molz, F.J. 1994. Variably saturated modeling of transient drainage: sensitivity to soil properties. Journal of Hydrology 161(1-4), 91-108.

Zhang, J., Wang, Z. and Luo, X. 2018. Parameter Estimation for Soil Water Retention Curve Using the Salp Swarm Algorithm. Water 10(6), 815.