

Article

Differential Evolution's Application to Estimation of Soil Water Retention Parameters

Zhonghui Ou

School of Mathematics and Computer Science, Fujian Normal University, 350117 Fuzhou, China;
E-Mail: zhou@fjnu.edu.cn; Tel.: +86-591-2286-8115.

Academic Editors: Ole Wendroth and Peter Langridge

Received: 11 May 2015 / Accepted: 14 September 2015 / Published: 18 September 2015

Abstract: A Differential Evolution (DE) is introduced to predict the parameters of the soil water retention curve (SWRC) and it is configured for reliability and efficiency with the Unsaturated Soil Hydraulic Property Database (UNSODA). The main investigated dataset is 235 samples from lab_drying_h-t table and the testing shows that the data resource is reliable and steady. Some specific statistical computations are designed to investigate the convergence speed and the fitness precision of DE, different measurements of hydraulic data, and parametric characteristics of textural groups. The statistical results on UNSODA show that DE has higher performance in parameter fitness and time saving than some previous optimization methods and the statistical values of soil water retention parameters (SWRP) can be directly applied in the agricultural research and practice.

Keywords: van Genuchten model; Differential Evolution; soil water retention parameter; UNSODA

1. Introduction

The Van Genuchten model (vG model) describes the soil-water content-pressure head curve and the closed-form relative hydraulic conductivity expression in unsaturated soils derived from the predictive conductivity models of Burdine or Mualem [1–5]. At the beginning, Genuchten numerically obtained the residual soil-water content θ_r ($\text{cm}^3\text{cm}^{-3}$), scaling parameter α (>0 , in cm^{-1}) inversely proportional to the air entry pressure, and the pore-size distribution index n (>1) by the semi-analytical and semi-graphical method and the nonlinear least-squares curve-fitting method while the saturated

soil-water content θ_s and the saturated hydraulic conductivity K_s were measured experimentally [3]. The unsaturated hydraulic conductivity K_r was predicted well in the cases of Hygiene sandstone, Touchet Silt Loam G.E. 3, Silt Loam G.E.3 and Guelph Loam, but Beit Netofa Clay was not well predicted [3].

It is difficult to estimate SWRPs because of the spatial and temporal variability of the soil hydraulic properties in the field. In addition, pedotransfer functions (PTF) help to convert the directly-measured data from soil survey (e.g., field morphology, soil texture, structure, pH, *etc.*) into estimates of soil hydraulic parameters [6–9]. Five hierarchical neural network PTFs fulfilled this conversion in vG model from different levels of input data in Unsaturated Soil Hydraulic Property Database (UNSODA), and the neural network analyses combining with the bootstrap method generated uncertainty estimates of the predicted hydraulic properties and statistically appraised the reliability of the predictions [10–15].

Some quasi-physical models have estimated the SWRC according to the shape similarity between SWRC and the cumulative particle-size distribution [16–18].

The fractal geometry model can mathematically and physically describe porous media and build functional relationships between the water content and the matric potential. It has three SWRC types based on the fractal organization of soil structure: fractal mass, fractal surface and fractal pore-size distribution [19–22]. Ghanbarian *et al.* [23] used a relationship between the pore size distribution index of the Brooks and Corey model (BC model) and the fractal dimension of SWRC to evaluate two approaches for estimating parameter m in vG model [3] and the statistical parameters showed that the approach proposed by Lenhard *et al.* [24] provided better estimates of m .

The predictions of parameters in SWRC are severely restricted by the experiment method and data. The accurate identification of SWRP (θ_r , α and n) and hydraulic conductivity made demand on cumulative outflow as input data in the one-step pressure outflow experiments and the initial parameter evaluation to be reasonably close to their true values [25,26]. An integral method with Richards' equation and the closed-form equations of soil hydraulic properties was used to estimate α and n if both infiltration and wetting front with time in a horizontal absorption experiment were recorded and it then provided a transient water flow approach to estimate SWRC instead of the usual equilibrium method [27,28].

UNSODA is a database with unsaturated soil hydraulic properties and other soil information [14]. UNSODA version 1 consists of 791 and version 2 of 790 soil materials with water retention, saturated and unsaturated hydraulic conductivity data measured in the field or laboratory, as well as particle size distribution and bulk density data. Each soil material has an identifier code, the minimum is 1010, the maximum is 4960, the soil material with same identifier code in one table is defined as a sample, and the data size of a sample is the number of data records with the same identifier code in one table. The data of each soil material are classified to different tables as a consecutive series of records with the same identifier code according to three hierarchic levels: measurement methodology (field or lab); hydraulic drainage curve (drying or wetting); data relationship (preshead-conductivity, preshead- θ , θ -diffusivity or θ -conductivity). For instance, there are 700 soil samples with available preshead- θ paired data and 90 samples missing data in the lab_drying_h-t table. Kosugi developed a general conductivity model for soils with lognormal pore-size distribution based on the Mualem-Dagan pore-scale model and two predictive methods reducing the average prediction error more than 77% compared with the Burdine and Mualem predictive models with use of 200 soil samples in UNSODA [29,30]. UNSODA was the database of Neural network analysis, bootstrap method and ROSETTA model implementing five

PTFs for hierarchical estimation of the soil water retention and the saturated or unsaturated hydraulic conductivity [10–12].

DE invented by Kenneth Price and Rainer Storn is a very simple population based, stochastic function minimizer for continuous function optimization and optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality [31–33]. The crucial idea behind DE is a scheme for generating trial parameter vectors. Basically, DE adds the weighted difference between two population vectors to a third vector. DE optimizes a problem by maintaining a population of candidate solutions and creating new candidate solutions by combining existing ones according to its simple formulae, and then keeping whichever candidate solution has the best score or fitness on the optimization problem at hand. It has become a popular optimization method widely used in Informatics, Thermodynamics, dynamic systems, etc., and it is already integrated in Mathematica and MATLAB. At present, we mainly focus on the compatibility and adaptation of DE and UNSODA.

In this paper, we will show how to generate different datasets from UNSODA, estimate the SWRPs in vG model on each dataset, and obtain more reliable statistical results with DE. First we will correctly configure and test DE with UNSODA; then we will design statistical computations to estimate SWRPs and compare with previous algorithms and results; finally we will present the SWRP tables which can be referred directly by agricultural research and practice.

2. Differential Evolution

DE is a parallel direct search method which utilizes NP D-dimensional parameter vectors $\mathbf{x}_{i,G}$ ($i = 1, 2, \dots, NP$) as a population for each generation G and it consists of 4 basic steps: initialization, mutation, crossover and selection [31,33]. N_p is the number of parameters to be optimized.

2.1. Initialization

The initial vector population is randomly selected and should cover the entire parameter space assuming a uniform probability distribution for all random decisions. If a preliminary solution is ready, the initial population might be generated by adding normally distributed random deviations to the nominal solution,

$$\mathbf{x}_{i,G} = \mathbf{x}_{min} + \text{rand}(D)(\mathbf{x}_{max} - \mathbf{x}_{min}), \quad G = 0 \quad (1)$$

where D is the dimension of parameter vector, and \mathbf{x}_{max} and \mathbf{x}_{min} are the lower and upper bounds of the parameter vectors $\mathbf{x}_{i,G}$.

2.2. Mutation

DE adds the weighted difference between two population vectors to a third one in order to generate new parameter vectors. For each target vector $\mathbf{x}_{i,G}$ ($i = 1, 2, \dots, NP$), a mutant vector is generated by

$$\mathbf{v}_{i,G+1} = \mathbf{x}_{r_1,G} + F(\mathbf{x}_{r_2,G} - \mathbf{x}_{r_3,G}) \quad (2)$$

where the mutually different random indexes $r_1, r_2, r_3 \in \{i = 1, 2, \dots, NP\}$ are different from the running index i , and $F \in [0, 2]$ is a constant factor which controls the amplification of the differential variation $(\mathbf{x}_{r_2,G} - \mathbf{x}_{r_3,G})$.

2.3. Crossover

Crossover is to increase the diversity of the perturbed parameter vectors. The mutated vector's parameters are mixed with the parameters of another predetermined vector, *i.e.*, the target vector, to yield a trial vector

$$\mathbf{u}_{i,G+1} = (\mathbf{u}_{1i,G+1}, \mathbf{u}_{2i,G+1}, \dots, \mathbf{u}_{Di,G+1}) \quad (3)$$

where

$$\mathbf{u}_{ji,G+1} = \begin{cases} \mathbf{v}_{ji,G+1} & \text{if } (\text{randb}(j) \leq \text{CR}) \text{ or } j = \text{rnbr}(i) \\ \mathbf{x}_{ji,G} & \text{if } (\text{randb}(j) > \text{CR}) \text{ and } j \neq \text{rnbr}(i) \end{cases} \quad (4)$$

$j = 1, 2, \dots, D$. In Equation (4), $\text{randb}(j)$ is the j th evaluation of a uniform random number generator within the range $[0, 1]$, CR is the crossover constant in $[0, 1]$ determined by the investigated problem, and $\text{rnbr}(i)$ is the randomly chosen index in $[1, 2, \dots, D]$ ensuring that $\mathbf{u}_{i,G+1}$ gets at least one parameter from $\mathbf{v}_{i,G+1}$.

2.4. Selection

To decide whether or not it should become a member in generation $G + 1$, the trial vector $\mathbf{u}_{i,G+1}$ is compared to the target vector $\mathbf{x}_{i,G}$ using the greedy criterion. If $\mathbf{u}_{i,G+1}$ yields a smaller cost function value than $\mathbf{x}_{i,G}$, $\mathbf{x}_{i,G+1}$ is set to $\mathbf{u}_{i,G+1}$ in the next generation $G + 1$; otherwise, the old value $\mathbf{x}_{i,G}$ is retained to $\mathbf{x}_{i,G+1}$.

Moreover, the iteration termination conditions of the main algorithm in DE are relevant to the fitness effect and the convergence precision, and we adopt some usual methods: Maximum Evolution Generation (MEG), Maximum Number of Iterations (MIT) or some given objective value (GOV) for RMSE_w .

During the test and selection of algorithms, we found that DE can get higher precision and faster speed mainly depends on such characteristics: random initialization without predetermined initial parameter values, vector-based computation, global optimization and evolution strategy. Some algorithms, *e.g.*, neural network, support vector regression, genetic algorithms, do not include all these characteristics or are not easily implemented in pSWRP estimation [6–9,19–22,35–37].

3. Fitting Soil Water Retention Parameters to Hydraulic Data

The soil water retention function in vG model is given by

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha h)^n]^m} \quad (5)$$

where $m = 1 - 1/n$, and $\theta(h)$ is the volumetric water content ($\text{cm}^3\text{cm}^{-3}$) at the pressure head h (cm, taken positive). The dimensionless effective water content is

$$S_e = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} \quad (6)$$

The objective function that is minimized for fitting SWRC, *i.e.*, Equation (5) to the prehead- θ data in UNSODA and then gets the optimized SWRPs is

$$O_w(\mathbf{p}) = \sum_{i=1}^{N_w} (\theta_i - \theta'_i)^2 \quad (7)$$

where θ_i and θ'_i are the measured and the estimated water contents respectively, N_w is the number of measured water retention points for each sample and is taken as the data size of a sample in computation, and \mathbf{p} is the parameter vector $(\theta_r, \theta_s, \alpha, n)$. θ_i is from UNSODA, and θ_s and θ'_i will be calculated by DE. The goodness of fit of Equation (5) is quantified with the root mean square error:

$$RMSE_w(\mathbf{p}) = \sqrt{\frac{O_w(\mathbf{p})}{N_w - N_p}} \quad (8)$$

N_p is the number of parameters to be optimized. Results of Equation (8) will be presented as averages for each textural group or for the investigated dataset and also be taken as an iteration termination condition in programming [12].

Before we apply DE to estimate SWRPs, it is necessary to configure DE correctly: choose an appropriate dataset from UNSODA for specific task, balance the convergency and time consumption, and evaluate the control variables.

- The choice of dataset. There are four tables about preshead- θ data in UNSODA: field_drying_h-t, field_wetting_h-t, lab_drying_h-t and lab_wetting_h-t. There are 127 samples in field_drying_h-t table, 0 samples in field_wetting_h-t table, 700 samples in lab_drying_h-t table, and 28 samples in lab_wetting_h-t table with available data and this dataset is called dataset 1. The data in dataset 1 are diverse and heterogeneous (field or lab, wetting or drying), and appropriate for looking into the universal characteristics of UNSODA and the feasibility and robustness of DE. Schaap and Leij [12] and Schaap *et al.* [13] chose 235 codes in lab_drying_h-t table as a dataset based on the criteria: quality of the data; presence of sufficient texture data; the data from the laboratory drying (drainage) branches; eliminating samples with low bulk density values ($<0.5 \text{ g/cm}^3$), and it is called dataset 2 here. Moreover, the designated 235 identifier codes with available data also partially appear in other two h-t tables: 26 samples in field_drying_h-t table and 1 sample in lab_wetting_h-t, the sizes of which are not enough for statistics. Tables 1 and 2 show: The $RMSE_w$ mean of dataset 1 becomes numerically stable at MEG = 400 and dataset 2 at MEG = 200; the $RMSE_w$ maximum of dataset 1 is 14 times of the mean, and only 3 times for dataset 2; the discrepancy between $RMSE_w$ maximum and minimum of dataset 1 is 3 times of dataset 2; the $RMSE_w$ mean of dataset 2 is 20% higher than dataset 1; the means of loop times are different only at MEG $\geq 10,000$; the minimum loop times of dataset 2 is 4~5 times of dataset 1. The above differences of $RMSE_w$ and loop times between datasets 1 and 2 lie in: Because the dataset 2 is

completely from lab_drying_h-t table and the data size of any sample in dataset 2 is not less than 6, the data quality will be higher and the statistical characteristics are more homogeneous.

- The iteration termination. DE is globally convergent, the convergence speed will become slower, $RMSE_w$ will not decrease for long time after reaching certain relatively steady value, and this value can be taken as a numerical convergence value (NCV). MEG or MIT shall satisfy that every sample in dataset can reach a NCV. The loop times that a sample takes to reach its own NCV is called the actual loop times and it should be less than MEG. The NCVs of $RMSE_w$ mean in Tables 1 and 2 can be respectively taken as 0.0106~0.0107 and 0.0121~0.0122 while means of loop times remarkably increase with MEG. In Table 2, $RMSE_w$ becomes stable at MEG = 200, but the corresponding mean loop times 197 is close to MEG and it can not ensure that some samples already get NCV when MEG and mean loop times are close. Therefore it is ideal to set MEG = 1000 or above as an iteration termination condition according to the first and the sixth columns in Table 2. If a sample reaches GOV, the iteration will terminate. Hence we utilize both MEG and GOV as the termination conditions in DE programming.
- The relationship between the loop times and the data size of a soil sample. The data size is the number of data records of a soil sample, *i.e.*, N_w in Equation (3). In Table 3, MEG = 10,000; in the minimum case of soil samples, the loop times is the minimum for all samples, and the maximum likewise; number is the count of the minimum or maximum cases; but the average data sizes, 13 and 12.23 of the minimum and the maximum cases are very close and it means that loop times are unrelated to data size; the maximum cases have much lower $RMSE_w$ mean and more loop times will get better fitness. It implies that DE time consumption depends on the convergency speed rather than the data size.
- The range of control variables. The rule of thumb values for the control variables in DE is: $F \in [0.5, 1.0]$, $Cr \in [0.8, 1.0]$ and $Np = 10 \cdot D$ [33]. The designated 235 identifier codes have only 26 samples with available data in the field_drying_h-t table, which composes dataset 3. We compare dataset 2 with 235 samples and dataset 3 with 26 samples by different values of Cr, 0.3 and 0.8 in Table 4 as MEG = 10,000: In dataset 2, different Crs do not generate evidently different estimates of water retention parameters; compared with dataset 2, the estimation differences in dataset 3 can not be ignored except θ_r and $RMSE_w$. It means the value of Cr is not crucial to the parametric estimation in UNSODA if the dataset is big enough and this conclusion can also be safely applied to other control variables and even the initial values of parameters.

After DE configuration and implement, we now apply it to predict the SWRPs of different soil texture groups and investigate its computation speed and goodness of fit. The composition of dataset 2 with 235 samples is: 112 sands, 37 loams, 55 silts and 31 clays. This composition is different from that used by Schaap and Leij [12], and Schaap *et al.* [13]. We think that UNSODA version 2 has been updated from version 1. We will examine this viewpoint by comparing the results in Table 5 with Table 1 in Schaap and Leij [12]: Most estimates of parameters θ_r , θ_s , $\lg(\alpha)$, $\lg(n)$ and even their standard deviations in the first row of two tables for two datasets are close and the highest relative difference is only 3%; however, θ_s of sands, $\lg(\alpha)$ of loams, θ_r and $\lg(n)$ of loams, silts and clays in both tables appear different and their relative differences are more than 12%. According to this comparison, we can confirm that the dataset 2 and the dataset used by Schaap and Leij [12] have selected the samples with the same 235 identifier

codes, but the compositions are different: some soil materials have been reclassified to another textural groups in UNSODA version 2. $RMSE_w$ s in the first 4 rows of Table 5 are almost 0.001 and is one order of magnitude lower than the values in Table 1 [12], which indicates SWRP estimates by DE are improved after grouping soil texture. The higher-precision prediction and fit of DE are also found in the comparison with nonlinear least-squares curve-fitting method or damper least square method [12]. Plant stresses (drought stress, flood stress) are related to θ_r and θ_s , and tensile and shear modulus and the constitutive variables of soil to $\lg(\alpha)$ and $\lg(n)$. Table 5 indicates that soil texture should be taken into the agriculture engineering and the precaution of geological disasters.

The first hierarchic level in UNSODA is the measurement methodology, field or lab. Dataset 3 has only 26 of 235 designated identifier codes with available data in field_drying_h-t table. The samples with the same 26 identifier codes in lab_drying_h-t table are called dataset 4. We can compare the field and lab measurements by dataset 3 and dataset 4 in Table 6: $RMSE_w$ of field is 10 times of the lab; the field and lab values of θ_r and $\lg(n)$ are evidently different, but θ_s and $\lg(\alpha)$ otherwise; the standard deviation of each entry in the lab row are much smaller than the field one. These results verify again that the data gotten from lab measurement are more steady and accurate.

Table 1. $RMSE_w$ and loop times of dataset 1.

MEG	$RMSE_w$				Loop Times			Time Consumption (h:m:s)
	Mean	Maximum	Minimum	Std. Deviation	Mean	Maximum	Minimum	
50	0.018	0.193	0.001	0.014	46	49	13	00:02:38
100	0.012	0.182	0.001	0.011	97	99	58	00:04:29
200	0.011	0.141	0.001	0.009	196	199	55	00:08:16
300	0.011	0.141	0.001	0.009	286	299	55	00:11:56
400	0.011	0.141	0.001	0.009	356	399	52	00:15:28
500	0.011	0.141	0.001	0.009	426	499	51	00:19:08
1000	0.011	0.141	0.001	0.009	748	999	50	00:37:11
10,000	0.011	0.141	0.001	0.009	3322	9999	49	07:51:09
40,000	0.011	0.141	0.001	0.009	10,247	39,999	57	27:15:55
Std. deviation	2.53×10^{-3}	2.07×10^{-2}	1.52×10^{-4}	1.44×10^{-3}				

Table 2. $RMSE_w$ and loop times of dataset 2.

MEG	$RMSE_w$				Loop Times		
	Mean	Maximum	Minimum	Std. Deviation	Mean	Maximum	Minimum
50	0.018	0.065	0.002	0.010	47	49	25
100	0.013	0.039	0.001	0.007	97	99	83
200	0.012	0.039	0.001	0.007	197	199	183
300	0.012	0.039	0.001	0.007	287	299	218
400	0.012	0.039	0.001	0.007	353	399	221
500	0.012	0.039	0.001	0.007	424	499	221
1000	0.012	0.039	0.001	0.007	722	999	254
10,000	0.012	0.039	0.001	0.007	2963	9999	240
40,000	0.012	0.039	0.001	0.007	7838	39,999	265
Std. deviation	2.067×10^{-3}	9.24×10^{-3}	3.26×10^{-4}	9.84×10^{-4}			

Table 3. Relationship between loop times and the data size of soil sample.

Case	Loop Times	Number	Data Size	RMSE _w Mean
minimum	240	1	13	0.017
maximum	9999	31	12.23	0.009

Table 4. CR range.

	Dataset 2		Dataset 3	
	Cr = 0.3	Cr = 0.8	Cr = 0.3	Cr = 0.8
θ_r	0.054	0.057	0.166	0.152
θ_s	0.463	0.466	0.442	0.476
$\lg(\alpha)$	−1.625	−1.607	−1.581	−1.495
$\lg(n)$	0.214	0.208	0.604	0.536
RMSE _w	0.001	0.001	0.015	0.015

Table 5. Average hydraulic parameters for each soil textural group with standard deviations in parentheses.

	N	θ_r (cm ³ cm ^{−3})	θ_s	$\lg(\alpha)$ (cm ^{−1})	$\lg(n)$	RMSE _w (cm ³ cm ^{−3})
All	235	0.057 (0.081)	0.467 (0.139)	−1.606 (0.550)	0.204 (0.197)	0.001 (0.002)
Sands	112	0.050 (0.044)	0.442 (0.154)	−1.534 (0.421)	0.321 (0.321)	0.001 (0.002)
Loams	37	0.096 (0.137)	0.525 (0.139)	−1.239 (0.488)	0.106 (0.109)	0.001 (0.002)
Silts	55	0.035 (0.064)	0.436 (0.097)	−1.890 (0.594)	0.106 (0.057)	0.002 (0.003)
Clays	31	0.071 (0.104)	0.542 (0.099)	−1.799 (0.644)	0.075 (0.051)	0.051 (0.001)

Sands: sand, loamy sand, sandy loam, sandy clay loam; Loams: loam, clay loam; Silts: silt loam, silt; Clays: clay, sandy clay, silty clay, silty clay loam.

Table 6. Comparison between field and lab measurements.

Measurement	θ_r	θ_s	$\lg(\alpha)$	$\lg(n)$	RMSE _w
field	0.155 (0.102)	0.454 (0.196)	−1.562 (0.557)	0.533 (0.354)	0.015 (0.064)
lab	0.075 (0.051)	0.494 (0.131)	−1.568 (0.311)	0.288 (0.186)	0.001 (0.003)

4. Results and Discussion

In Section 3, DE has already applied to estimate SWRPs after the dataset filtering, DE and UNSODA configuring and the statistical calculation compiling, and we have gotten some significant results from the calculation and Tables 1–6: The dataset composed of the designated 235 identifier codes in lab_drying_h-t table is more appropriate for statistical research and the results based on this dataset are credible; the data size of a soil sample has no direct relationship with the actual loop times in DE main iteration; the different values of control variables will not produce evident change on the average estimates of SWRPs if the dataset is big enough; the average RMSE_ws of dataset 2, sands, loams and silts groups are almost one order of magnitude lower than some previous optimization methods, and DE has higher precision and better fitness; the estimates of θ_s and $\lg(\alpha)$ from field and lab measurements are close, and θ_r and $\lg(n)$ otherwise; parameter estimates on the whole dataset 2 are close to the dataset used by Schaap and Leij [12], but there appear differences on the estimates of θ_r and $\lg(n)$ of loam, silt and clay groups; θ_r has showed different change trend from other parameters in statistical process.

The convergence speed is a fascinating issue when using database. It is already shown that there is no evident relationship between the loop times and the data size of a soil sample in Table 3. In Table 7: a and b represent two program runnings with the same MEG so that we can estimate the effect of stochastic factors in DE; the first column is actual loop times and the second is $RMSE_w$ under same MEG. Loop times change irregularly while $RMSE_w$ s maintain steady in a and b runnings under the same MEG; the sample with identifier code = 2763 has 13 pairs of data and gets the minimum loop times 240, and the sample with identifier code = 2660 has 12 pairs of data and gets the maximum loop times 19,999 when MEG = 20,000 a, b; in the row of code = 2763, $RMSE_w = 0.017$, but the actual loop times are extraordinarily higher when MEG = 5000 b and MEG = 20,000 a; in the row of code = 2660, the actual loop times fluctuate abruptly while $RMSE_w$ s swing in narrow range (0.012, 0.014), and the three smallest loop times 1117, 1982 and 1578 contrarily get the smallest $RMSE_w = 0.012790$ under MEG = 3000 a, 3000 b and 5000 b. These statistical results in Table 7 seem to contradict to the common sense: Similar $RMSE_w$ s or similar data sizes of a sample should take similar loop times; smaller $RMSE_w$ should take more loop times. We guess that the stochastic factors (e.g., random seed, mutation step, crossover step, etc.) in DE might have considerable impact on the computation speed and the optimization efficacy.

Table 7. Convergence speed.

Code	MEG															
	3000 a		3000 b		5000 a		5000 b		10,000 a		10,000 b		20,000 a		20,000 b	
2763	328	0.017	293	0.017	399	0.017	504	0.017	240	0.017	310	0.017	849	0.017	372	0.017
2660	1117	0.013	1982	0.013	4999	0.013	1578	0.013	9999	0.013	9998	0.013	19,999	0.014	19,999	0.013

5. Conclusions

In this paper we first introduced UNSODA and DE, and then applied DE to estimate SWRPs in vG model based on UNSODA. We have configured DE and tested specific datasets in order to derive statistically tenable results. We have discussed the relationship between the data size of a soil sample and the actual loop times in the DE main iteration for reaching a NCV. Some specific statistical computations have been designed to select appropriate datasets and compare the parameter estimates between different values of control variables, different measurements or different textual groups, etc., and the calculation results shows that DE is capable to derive the values of soil water retention parameters from UNSODA, UNSODA is a reliable database for soil hydraulic property indices and statistical Tables 1–7 can be applied directly in agricultural research and practice.

We have also statistically analysed the factors on convergence or fitness as the fundamental issues of DE. The methods above are illuminative for correctly using UNSODA and DE, and the conclusions are valuable for soil hydraulic parameters. In the future, We will predict SWRC and unsaturated hydraulic conductivity expression together with basic or advanced forms of DE, explore some fundamental issues in DE in order to explain some basic problems in the determination of soil hydraulic parameters and then propose new algorithmic, survey or experimental methods.

Acknowledgments

This work was financially supported by the Science and Research Startup Fund of Fujian Normal University, Nonlinear Analysis and Its Applications IRTL1206, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars from State Education Ministry DB-179.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Burdine, N.T. Relative permeability calculations from pore-size distribution data. *J. Pet. Technol.* **1953**, *5*, 71–78.
2. Mualem, Y. A new model predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.* **1976**, *12*, 513–522.
3. 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.
4. Van Genuchten, M.T.; Hopmans, J.W. A Decade of Multidisciplinary Research. *Vadose Zone J.* **2013**, *12*, doi:10.2136/vzj2013.08.0150.
5. Dettmann, U.; Bechtold, M.; Frahm, E.; Tiemeyer, B. On the applicability of unimodal and bimodal van Genuchten—Mualem based models to peat and other organic soils under evaporation conditions. *J. Hydrol.* **2014**, *515*, 103–115.
6. Rawls, W.J.; Pachepsky, Y.; Shen, M.H. Testing soil water retention estimation with the MUUF pedotransfer model using data from the southern United States. *J. Hydrol.* **2001**, *251*, 177–185.
7. Pachepsky, Y.A.; van Genuchten, M.T. Pedotransfer functions. *Encycl. Agrophysics* **2011**, 556–561.
8. Pan, F.; Pachepsky, Y.; Jacques, D.; Guber, A.; Hill, R.L. Data assimilation with soil water content sensors and pedotransfer functions in soil water flow modeling. *Soil Sci. Soc. Am. J.* **2012**, *76*, 829–844.
9. Ramos, T.B.; Gonçalves, M.C.; Brito, D.; Martins, J.C.; Pereira, L.S. Development of class pedotransfer functions for integrating water retention properties into Portuguese soil maps. *Soil Res.* **2013**, *51*, 262–277.
10. Schaap, M.G.; Leij, F.J.; van Genuchten, M.T. Neural Network Analysis for Hierarchical Prediction of Soil Hydraulic Properties. *Soil Sci. Soc. Am. J.* **1998**, *62*, 847–855.
11. Schaap, M.G.; Feike, J.L. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil Tillage Res.* **1998**, *47*, 37–42.
12. Schaap, M.G.; Leij, F.J. Improved Prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten Model. *Soil Sci. Soc. Am. J.* **2000**, *64*, 843–851.
13. 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.
14. Nemes, A.; Schaap, M.G.; Leij, F.J.; Wösten, J.H.M. Description of the unsaturated soil hydraulic database UNSODA version 2.0. *J. Hydrol.* **2001**, *251*, 151–162.

15. Pachepsky, Y.; Pan, F.; Martinez, G. Sensor Network Data Assimilation in Soil Water Flow Modeling. In *Application of Soil Physics in Environmental Analyses*; Springer International Publishing: Geneva, Switzerland: 2014; pp. 239–260.
16. Fredlund, M.D.; Wilson, G.W.; Fredlund, D.G. Use of the grain-size distribution for estimation of the soil-water characteristic curve. *Can. Geotech. J.* **2002**, *39*, 1103–1117.
17. Haverkamp, R.; Leij, F. J.; Fuentes, C.; Sciortino, A.; Ross, P. J. Soil water retention: I. Introduction of a shape index. *Soil Sci. Soc. Am. J.* **2005**, *69*, 1881–1890.
18. Leij, F.J.; Haverkamp, R.; Fuentes, C.; Zatarain, F.; Ross, P.J. Soil water retention: II. Derivation and application of shape index. *Soil Sci. Soc. Am. J.* **2005**, *69*, 1891–1901.
19. Bird, N.R.A.; Perrier, E.; Rieu, M. The water retention function for a model of soil structure with pore and solid fractal distributions. *Eur. J. Soil Sci.* **2000**, *51*, 55–63.
20. Perfect, E. Modeling the primary drainage curve of prefractal porous media. *Vadose Zone J.* **2005**, *4*, 959–966.
21. Huang, G.; Zhang, R. Evaluation of soil water retention curve with the pore-solid fractal model. *Geoderma* **2005**, *127*, 52–61.
22. Ghanbarian-Alavijeh, B. Modeling Physical and Hydraulic Properties of Disordered Porous Media: Applications from Percolation Theory and Fractal Geometry. Ph.D. Thesis, Wright State University, Dayton, OH, USA, 2014.
23. Ghanbarian-Alavijeh, B.; Liaghat, A.; Huang, G.H.; van Genuchten, M.T. Estimation of the van Genuchten soil water retention properties from soil textural data. *Pedosphere* **2010**, *20*, 456–465.
24. Lenhard, R.J.; Parker, J.C.; Mishra, S. On the correspondence between Brooks-Corey and van Genuchten models. *J. Irrig. Drain. Eng.* **1989**, *115*, 744–751.
25. Kool, J.B.; Parker, J.C.; van Genuchten, M.T. Determining soil hydraulic properties from one-step outflow experiments by parameter estimation: I. Theory and numerical studies. *Soil Sci. Soc. Am. J.* **1985**, *49*, 1348–1354.
26. Londra, P.A.; Valiantzas, J.D. Soil water diffusivity determination using a new two-point outflow method. *Soil Sci. Soc. Am. J.* **2011**, *75*, 1343–1346.
27. Shao, M.; Horton, R. Integral method for estimating soil hydraulic properties. *Soil Sci. Soc. Am. J.* **1998**, *62*, 585–592.
28. Peng, H.; Horton, R.; Lei, T.; Dai, Z.; Wang, X. A modified method for estimating fine and coarse fractal dimensions of soil particle size distributions based on laser diffraction analysis. *J. Soils Sediments* **2015**, *15*, 937–948.
29. Kosugi, K. Three-parameter lognormal distribution model for soil water retention. *Water Resour. Res.* **1994**, *30*, 891–901.
30. Kosugi, K. General model for unsaturated hydraulic conductivity for soils with lognormal pore-size distribution. *Soil Sci. Soc. Am. J.* **1999**, *63*, 270–277.
31. Storn, R.; Price, K. Minimizing the Real Functions of the ICEC'96 Contest by Differential Evolution. In Proceedings of the 1996 IEEE Conference on Evolutionary Computation, Nagoya, Japan, 20–22 May 1996; pp. 842–844.
32. César Trejo Zúñiga, E.; López Cruz, I.L.; García, A.R. Parameter estimation for crop growth model using evolutionary and bio-inspired algorithms. *Appl. Soft Comput.* **2014**, *23*, 474–482.

33. Storn, R. Differential Evolution Research—Trends and Open Questions. In *Advances in Differential Evolution*; Chakraborty, U.K., Ed.; Springer-Verlag: Heidelberg, Germany, 2008; pp. 1–32.
34. Van Genuchten, M.T.; Leij, F.J.; Yates, S.R. *The RETC Code for Quantifying the Hydraulic Functions of Unsaturated Soils*; EPA Report 600/2-91/065; US Salinity Laboratory, USDA, ARS: Riverside, CA, USA, 1991.
35. Garg, A.; Vijayaraghavan, V.; Wong, C.H.; Tai, K.; Sumithra, K.; Gao, L.; Singru, P.M. Combined ci-md approach in formulation of engineering moduli of single layer graphene sheet. *Simul. Model. Pract. Theory* **2014**, *48*, 93–111.
36. Garg, A.; Tai, K. Stepwise approach for the evolution of generalized genetic programming model in prediction of surface finish of the turning process. *Adv. Eng. Softw.* **2014**, *78*, 16–27.
37. Garg, A.; Garg, A.; Tai, K. A multi-gene genetic programming model for estimating stress-dependent soil water retention curves. *Comput. Geosci.* **2014**, *18*, 45–56.

© 2015 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).