Soil Water Retention Characteristics of Vertisols and Pedotransfer Functions Based on Nearest Neighbor and Neural Networks Approaches to Estimate AWC

N. G. Patil[1]; D. K. Pal[2]; C. Mandal[3]; and D. K. Mandal[4]

Abstract: Irrigation management in vertisols is one of the major challenges to increase agricultural productivity in India and many developing countries. Unfortunately, information on hydraulic properties of these soils is very sparse. In an attempt to understand these soils for better management, 10 different functions were evaluated for their efficacy to describe soil-water retention characteristics (SWRC) of vertisols of India, and point pedotransfer functions (PTFs) were developed by using a nearest neighbor (*k*-NN) algorithm as an alternative to widely used artificial neural networks (ANN) for prediction of available water capacity (AWC). Soil profile information of 26 representative sites comprising 157 soil samples was used for analysis. The Campbell model fit to measured SWRC data better than any other model, with relatively lower root mean square error (RMSE) (0.0199), higher degree of agreement (0.9867), and lower absolute error on an average (0.0134). Three other functions, namely, modified Cass-Hutson, Brooks-Corey, and van Genuchten, also described the SWRC data with acceptable accuracy. Four levels of input information were used for point pedotransfer function (PTF) development: (1) textural data [data on sand, silt, and clay fraction (SSC)]; (2) Level 1 + bulk density data (SSCBD); (3) Level 2 + organic matter (SSCBDOM); and (4) Level 1 + organic matter (SSCOM). The RMSE in predictions by *k*-NN PTFs ranged from 0.0339 to 0.0450 $m^3\,m^{-3}$ with an average of 0.0403 $m^3\,m^{-3}$. The ANN PTFs performed with average RMSE 0.0426 $m^3\,m^{-3}$ and a range of 0.0395 to 0.0474 $m^3\,m^{-3}$. The *k*-NN algorithm provided a viable alternative to neural regression with marginally better performance and the benefit of flexibility in the appending reference database. The results are significant because SWRC data are still in the development stage in India, and *k*-NN PTFs would have a greater value because of the flexibility. **DOI: 10.1061/(ASCE)IR.1943-4774.0000375.** *© 2012 American Society of Civil Engineers.*

CE Database subject headings: Soil water; Irrigation; India; Developing countries; Neural networks.

Author keywords: Pedotransfer functions; Neural networks; K nearest neighbor; van Genuchten function; Vertisols; Available water capacity.

Introduction

Vertisols (clay soils) have great potential for agricultural production, but many, especially in the developing world, are underutilized due to a lack of understanding regarding their behavior and management. In water-use regulation or budgeting, there are special problems associated with these soils compared to other soils. The problems are associated with two main factors, i.e., what constitutes available water capacity of the soils and rooting depth (Ahmad and Mermut 1996). Vertisols and vertic intergrades cover an area of 257×10^6 ha globally, and India's share is stated to be around 30% (Dudal 1965). The swell-shrink nature of these soils leads to complex hydraulic behavior, causing difficulties in managing them. In fact, management of these soils is one of the major challenges in increasing agricultural productivity in India. In developing countries like India, large-scale laboratory data on soil-water retention characteristics (SWRC) are rarely available primarily because of a lack of facilities and the costs involved. Because understanding soil-hydraulic properties is a prerequisite for any irrigation planning or hydrologic simulation, a lack of information on SWRC is a major constraint faced in India. Literature provides many mathematical functions to describe SWRC. Researchers have pursued a universal function that can describe SWRC of all types of soils. But experience shows that no single function can be termed generic, although the van Genuchten (VG) function historically has been the most widely adopted (Chang et al. 2004).

An indirect estimation of soil hydraulic properties (e.g., SWRC) has received attention from many researchers as an alternative to direct measurement/estimation. Basic soil information is often related empirically to properties of interest, mostly using regression tools. Most of the PTFs reported in the literature pertain to the estimation of SWRC. The PTF could be built to predict a point of interest on the SWRC curve, such as field capacity (soil-water retained at −33 kPa), permanent wilting point (soil-water retained at −1,500 kPa), and subsequently available water capacity (AWC). Point PTFs are argued to be better than parametric PTFs because soil-water retention in different ranges of soil-water potential is affected by different basic soil properties. However, parametric PTFs offer an advantage of continuous simulations (soil-water retention at any level of potential can be predicted) and facilitating prediction

[1]Scientist (SG) National Bureau of Soil Survey and Land Use Planning (NBSS&LUP), Amravati Road, Shanakarnagar P.O. Nagpur, India 440010 (corresponding author). E-mail: nitpat03@yahoo.co.uk

[2]Head, Division of SRS, National Bureau of Soil Survey and Land Use Planning (NBSS&LUP), Amravati Road, Shanakarnagar P.O. Nagpur, India 440010.

[3]Pr. Scientist, National Bureau of Soil Survey and Land Use Planning (NBSS&LUP), Amravati Road, Shanakarnagar P.O. Nagpur, India 440010.

[4]Pr. Scientist, National Bureau of Soil Survey and Land Use Planning (NBSS&LUP), Amravati Road, Shanakarnagar P.O. Nagpur, India 440010.

Note. This manuscript was submitted on June 8, 2010; approved on April 28, 2011; published online on April 30, 2011. Discussion period open until July 1, 2012; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Irrigation and Drainage Engineering***, Vol. 138, No. 2, February 1, 2012. ©ASCE, ISSN 0733-9437/2012/2-177–184/$25.00.

J. Irrig. Drain Eng. 2012.138:177-184.

of hydraulic conductivity at varied soil moisture levels. In applications such as irrigation decisions (scheduling criteria) or estimating crop water requirements, information on two points on the soil-water retention curve, namely, field capacity (FC) and permanent wilting point (PWP) or AWC, may suffice. For such straight applications uncomplicated point PTFs are desirable.

Most of the recent PTFs reported in the literature have used a neural regression approach (e.g., Jain et al. 2004; Minasny et al. 1999; Minasny and Mc Bratney 2002; Patil et al. 2010; Schaap et al. 1998). The data used for calibrating PTFs should account for most of the variations that are likely to be encountered in the soilscape of the area where they are meant to be used. Because there is no well-defined or developed soil database in India, it is imperative that any effort to develop PTFs should also consider future developments and that provisions be made for modifying PTFs. Unfortunately, regression PTFs do not provide such flexibility. This is perhaps the main reason why not a single report from India is found in the literature that makes use of flexible/alternative pattern-recognition algorithms. Nemes et al. (2006a, 2006b) have reported efficacy of the k-nearest neighbor (k-NN) approach to predict FC and PWP. They found that k-NN PTFs were as efficient as the PTFs developed using the most advanced neural computing techniques. These results are very important from an Indian context.

To the best of our knowledge, vertisols have not been studied for their soil-water retention characteristics (SWRC) on regional scale in India and information on parametric function that describes SWRC of these soils is not known. We hypothesized that point PTFs based on native data should be of greater value than the generic PTFs, and any PTF that provides flexibility of appending reference databases and acceptable predictive ability would be an asset for future development. The objectives of this study were to (1) identify a parametric function that best describes soil-water retention characteristics of vertisols; (2) establish point PTFs for the estimation of AWC using neural regression and the k-NN technique; (3) compare estimates by different PTFs.

Materials and Methods

We used a database developed by Pal et al. (unpublished report, 2003) that included basic soil information and soil-water retention properties. The database contains information on 26 profiles collected from the Indian states of Madhya Pradesh, Maharashtra, Karnataka, Andhra Pradesh, Tamil Nadu, Gujarat, and Rajasthan (Fig. 1). The profiles represent subhumid (moist), subhumid (dry), semiarid (moist), semiarid (dry), and arid climatic regions with mean annual rainfall (MAR) of 1,448–1,127 mm, 1,084–1,011 mm, 977–924 mm, 842–583 mm, and < 533 mm, respectively. The majority of these soils were developed in the alluvium of the weathered Deccan basalt. Undisturbed soil blocks (8 cm long, 6 cm wide, and 5 cm thick) were collected from soil horizons, and thin sections were prepared by the methods of Jongerius and Heintzberger (1975). The particle-size distribution of the collected soil samples was determined by the international pipette method after removal of organic matter. Sand (2,000–50 μm), silt (50–2 μm), total clay (< 2 μm), and fine clay (< 0.2 μm) fractions were separated according to the procedure of Jackson (1973). A seven-point SWRC (-33, -100, -300, -500, -800, $-1{,}000$, and $-1{,}500$ kPa) was mapped using pressure-plate apparatus. The sieved soil sample(s) were filled in rubber soil retainer rings of 6 cm diameter and 1 cm height on a ceramic plate of requisite capacity. The soil in the ring was allowed to saturate for 24 h with an excess of water, and the predetermined pressure from a source of compressed air was applied the next day. Moisture was determined

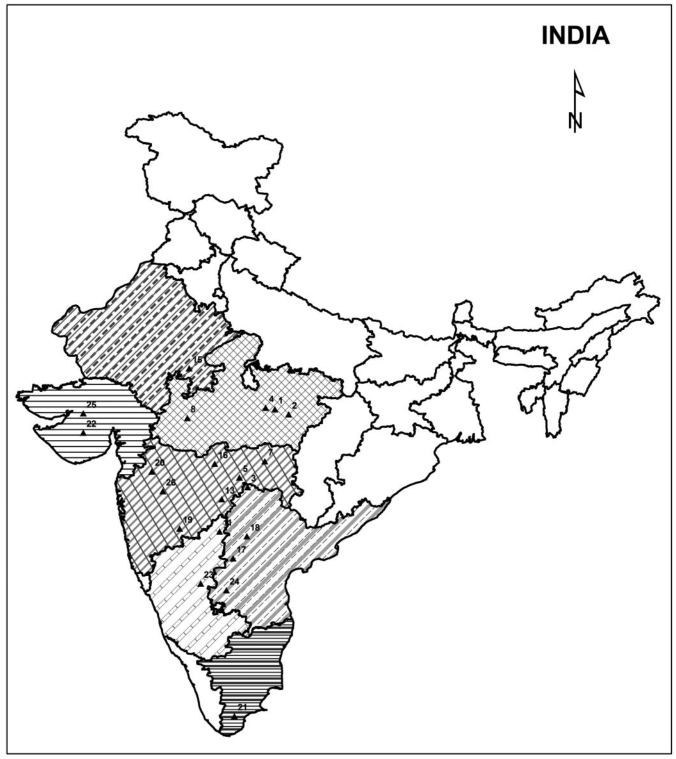

Fig. 1. Location of vertisol profiles in different states of India

gravimetrically after the soils had attained equilibrium at particular pressure. Other soil properties (not so important in the context of this article) were determined using standard soil-survey methodology and laboratory techniques.

Salient features of the reference database used in the study are presented in Table 1. Except for sand content, all the basic soil properties exhibited a relatively lower coefficient of variation (COV), implying lesser spatial changes. Seven-point data on soil-water retention characteristics for 157 samples were used for analysis. Thus, a total of 1,099 paired data on soil-water retained at varied suction pressure kPa were used. Salient statistical features of the SWRC data are presented in Table 1. Water retention ranged from 0.081–0.576 $m^3\,m^{-3}$. The magnitude of the COV at different points suggested that there was consistency in retention values. Mean standard deviation (SD) (measured data) was 0.006. We considered this SD value as criteria for judging the suitability of evaluated SWRC functions.

SWRC data were fitted to the parametric relationship between water content, θ, and the water potential of the soil, h, as described by different researchers. A power law equation suggested by Brooks and Corey (1966) describes this relationship as

$$S = (h_b/h)^{\lambda} \quad \text{for } h < h_b \qquad S = 1 \quad \text{for } h > h_b \tag{1}$$

where S is the saturation degree

$$S = (\theta - \theta r)/(\theta s - \theta r) \tag{2}$$

where θ = water content at pressure h; θs = maximum water content; θr = residual water content; h_b = air entry pressure head; and λ = pore distribution index.

Another most widely used function suggested by van Genuchten (1980) describes the relationship as

$$S = 1/[1 + (\alpha * h)^{n}]^{m} \tag{3}$$

J. Irrig. Drain Eng. 2012.138:177-184.

Table 1. Statistical Summary of Soil Properties of 157 Soil Samples

Property	Mean	SE	Variance	COV	Minimum	Maximum
Sand (%)	0.1	0.01	0.01	1.16	0	0.49
Silt (%)	0.33	0.01	0.01	0.26	0.16	0.52
Clay (%)	0.57	0.01	0.01	0.21	0.12	0.79
Bulk density	1.46	0.01	0.02	0.09	1.1	1.8
Organic matter	0.52	0.02	0.06	0.46	0.08	1.55
FC	0.38	0.01	0.01	0.21	0.21	0.58
PWP	0.2	0	0	0.24	0.08	0.32
Soil-water retention at suction pressure (-kPa)						
33	0.3835	0.0066	0.0068	0.2143	0.214	0.576
100	0.3279	0.0057	0.0051	0.2187	0.157	0.493
300	0.2699	0.0043	0.0029	0.1997	0.13	0.375
500	0.2552	0.0043	0.0029	0.2126	0.119	0.36
800	0.2388	0.0041	0.0026	0.2126	0.116	0.352
1000	0.223	0.0039	0.0024	0.2197	0.109	0.343
1500	0.1951	0.0037	0.0022	0.2383	0.081	0.323

Note: standard error (SE).

This equation is mostly under the assumption of $m = 1 - 1/n$. The value of α is related to the inverse of the air entry suction, $\alpha > 0$, and n is a measure of the pore-size distribution, $n > 1$.

Campbell (1974) described the water-retention function as

$$\theta = \theta_s (h/h_b)^{-1/b} \quad \text{for } h < h_b \tag{4}$$

$$\theta = \theta_s \quad \text{for } h \geq h_b \tag{5}$$

Hutson and Cass (1987) modified the Campbell function known as Cass-Hutson (CH) function

$$\theta = \theta_s \left(\frac{h}{a}\right)^{\frac{-1}{b}} \quad \text{for } \theta < \theta_i \tag{6}$$

$$\theta = \theta_s - \left[\theta_s h^2 \frac{(1 - \theta_i/\theta_s)}{a^2(\theta_i/\theta_s)}\right] \quad \text{for } \theta \geq \theta_i \tag{7}$$

Here,

$$\theta_i = \frac{2b\theta_s}{1 + 2b}$$

where a, b = empirical parameters; and h_b = air entry pressure.

Other functions evaluated in this study are enlisted in Table 2. A public domain computer code SWRC was used for fitting water-retention functions.

Table 2. Functions Selected to Describe SWRC of Study Soils

Function	Analytic expression
Matric potential as dependent variable	
Exponential	$h = \alpha e{-}^{\beta\theta}$
Power	$h = \alpha\theta^{-\beta}$
Farrel and Larson (1972)	$h = h_b e^{\alpha[1-(\theta-\theta r/\theta s-\theta r)]}$
Simmons et al. (1979)	$h = \alpha[^{e\beta(\theta-\Phi)} - 1]$
Libardi et al. (1979)	$h = \alpha[e^{\beta(\theta-\theta s)} - 1]$
Soil-water content as a dependent variable	
Driessen (1986)	$\theta = \theta_s h^{-r\ln(h)}$

Note: α, β, Φ, r are function parameters.

Two techniques were used to build the PTF, namely, artificial neural networks (ANN)-based regression and k-NN. Software developed by Nemes et al. (2008) to build PTFs for estimating field capacity (FC) and a permanent wilting point (PWP) from basic soil properties like textural distribution, bulk density, and organic matter in hierarchical order was used for building k-NN PTFs. The software/tool combines the k-NN algorithm with the bootstrap data-subset selection technique to allow the development of model ensembles that can be used to estimate the uncertainty of the final model output. For developing ANN-based PTFs, software called Neurointelligence was used. On the basis of earlier experience (Patil et al. 2010), a feed-forward neural-network model with three hidden nodes was preferred. The data set was partitioned into training (117 samples) and test (22 samples) sets. Upon finding an appropriate network model, the PTF was calibrated. For network training, the Levenberg-Marquardt algorithm was chosen because the data were small. Four levels of input information were used to avoid possible bias toward one set of inputs, and dependencies between basic soil properties and FC/PWP were established:

- Input Level 1: textural data [data on sand, silt, and clay fraction (SSC)];
- Input Level 2: Level 1 + bulk density data (SSCBD);
- Input Level 3: Level 2 + organic matter (SSCBDOM); and
- Input Level 4: Level 1 + organic matter (SSCOM).

Performance Evaluation

The efficacy of parametric functions was evaluated based on (1) root mean square error (RMSE), (2) index of agreement (d), (3) maximum absolute error (ME), (4) mean absolute error (MAE), and (5) correlation coefficient (r) and coefficient of determination (R^2). Statistics for RMSE, d, ME, and MAE were calculated by using the following equations, respectively, where n represents the number of data used for modeling and M_i and E_i represent measured and computed value, respectively, and the unit of errors is $m^3\ m^{-3}$:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(Ei - Mi)^2}{n}} \tag{8}$$

$$d = 1 - \frac{\sum_{i=1}^{n}(Ei - Mi)^2}{\sum_{i=1}^{n}(|Ei - \bar{M}| + |Mi - \bar{M}|)^2} \tag{9}$$

J. Irrig. Drain Eng. 2012.138:177-184.

Table 3. Statistical Indices to Judge Efficacy of Different Soil-Water Retention Functions in Describing SWRC Data of Vertisols

Index/Function	Driessen	Exponential	Farrel-Larson	LRN	Power	SNB	VG	CH	Campbell	BC
RMSE	0.118	0.102	0.1025	0.1081	0.1022	0.1019	0.0267	0.0213	0.0199	0.0302
d	0.5071	0.5767	0.5752	0.5403	0.5712	0.5782	0.9703	0.9849	0.9867	0.9599
ME	0.3570	0.3110	0.3170	0.3400	0.3090	0.3140	0.1818	0.1342	0.1331	0.0189
MAE	0.08989	0.0844	0.0848	0.088	0.0843	0.0843	0.0157	0.0144	0.0134	0.0184
R^2	0.6800	0.9217	0.9512	0.8765	0.9391	0.949	0.9140	0.651	0.9512	0.8970
				Clay samples (111)						
(-kPa)	−33	−100	−300	−500	−800	−1,000	−1,500	Mean	Function	
RMSE	0.0668	0.0422	0.0147	0.0132	0.0113	0.0143	0.0346	0.0282	Brooks-Corey	
d	0.74	0.86	0.98	0.98	0.98	0.98	0.86	0.91		
ME	0.1897	0.1121	0.0588	0.0346	0.0322	0.0334	0.0895	0.0786		
MAE	0.0484	0.0318	0.0111	0.0105	0.0090	0.0113	0.0269	0.0213		
RMSE	0.0384	0.0147	0.0121	0.0149	0.0170	0.0170	0.0151	0.0185	Campbell	
d	0.95	0.99	0.99	0.98	0.97	0.97	0.97	0.97		
ME	0.1331	0.0384	0.0388	0.0433	0.0484	0.0484	0.0374	0.0554		
MAE	0.0267	0.0117	0.0095	0.0120	0.0131	0.0131	0.0115	0.0139		
RMSE	0.0431	0.0161	0.0142	0.0175	0.0183	0.0194	0.0156	0.0206	Cass-Hutson	
d	0.93	0.98	0.98	0.97	0.96	0.95	0.96	0.96		
ME	0.1342	0.0440	0.0462	0.0421	0.0484	0.0695	0.0404	0.0607		
MAE	0.0274	0.0129	0.0117	0.0147	0.0140	0.0135	0.0129	0.0153		
RMSE	0.0636	0.0324	0.0145	0.0121	0.0125	0.0118	0.0252	0.0246	van Genuchten	
d	0.77	0.92	0.98	0.98	0.98	0.98	0.92	0.93		
ME	0.1818	0.0226	0.0466	0.0347	0.0297	0.0334	0.0690	0.0597		
MAE	0.0452	6.0206	0.0113	0.0095	0.0101	0.0091	0.0185	0.8749		

Note: Detailed indices for four screened functions and only mean indices for rest six functions are presented; maximum absolute error (ME); mean absolute error (MAE); coefficient of determination (R^2); Campbell-Hutson (CH), Brooks-Corey (BC); van Genuchten (VG); Libardi, Reichardt, and Nascimento (LRN); Simmons, Nielsen, and Biggar (SNB).

$$\mathrm{ME} = \mathrm{Max}|Ei - Mi| \tag{10}$$

$$\mathrm{MAE} = \sum_{i=1}^{n} \frac{|Ei - Mi|}{n} \tag{11}$$

$$\text{Linear correlation coefficient } r = \frac{1}{n-1} \sum_{i=1}^{n} \frac{(M_i - \bar{M})(E_i - \bar{E})}{S_M S_E} \tag{12}$$

Performance of the k-NN algorithm was evaluated against estimations made by neural network models, developed using the same data and input soil attributes. The same set of statistical indices was used for the comparison of measured and estimated data. The RMSE statistic indicates the model's ability to predict away from the mean. The value of the RMSE imparts more weight-to-high values because it involves the square of the difference between observed and predicted values. Ideally, the model should have the smallest MAE and smallest overall dispersion (RMSE). Units of all the errors in this paper are $m^3\,m^{-3}$.

Results and Discussions

Evaluation of SWRC Functions

The performance of 10 different models in describing SWRC could be judged from the statistical indices (Table 3). It was apparent that the Campbell model fitted better than any other model as evidenced by a relatively lower RMSE (0.0199), a highest degree of agreement (0.9867), and a mean absolute error (0.0134). Barring maximum error, which was lower in the BC model, all other indices clearly pointed to the superiority of the Campbell function in describing SWRC data of the black soils. The next best was the modified Cass-Hutson function in terms of performance, followed by VG and BC. All other functions were considered poor because RMSE exceeded 0.1 as against RMSE of 0.006 in the measured SWRC data. Although the van Genuchten model historically has been the most widely adopted (Chang et al. 2004), it was a third choice after the Campbell model and its variant (CH). Nevertheless, all the indices suggested that any of the four functions, namely, VG, BC, Campbell, and CH, could be used for fitting SWRC data. The BC and VG models are derived on a similar philosophical basis—both models specify the soil-water retention curve on an empirical basis with the soil hydraulic conductivity as a function of soil-water content. The Campbell and CH models can be considered as special cases of BC or simply power law. The soil-hydraulic functions of the VG model are more difficult to calculate and are considered relatively difficult to achieve a rapid numerical solution to the Richards' equation. The BC function does not fit relatively well to observed data in certain fine-textured soils, which was confirmed through the reported findings. Wagner et al. (1998) opine that the Brooks-Corey model perhaps represents capillarity better than adsorption. Different modifications to the BC function have been made by researchers to improve description near saturation. However, our findings show that the dry range could not be described well enough by the BC function.

Graphical representation (Fig. 2) shows that the Campbell model tended to overestimate the soil-water retention in a higher suction range (−33 and −100 kPa). The CH model also showed

J. Irrig. Drain Eng. 2012.138:177-184.

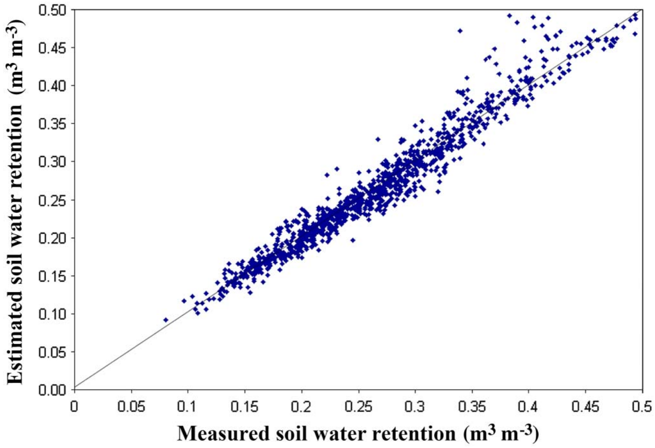

Fig. 2. Measured and estimated soil-water retention described by the Campbell function (best fitting)

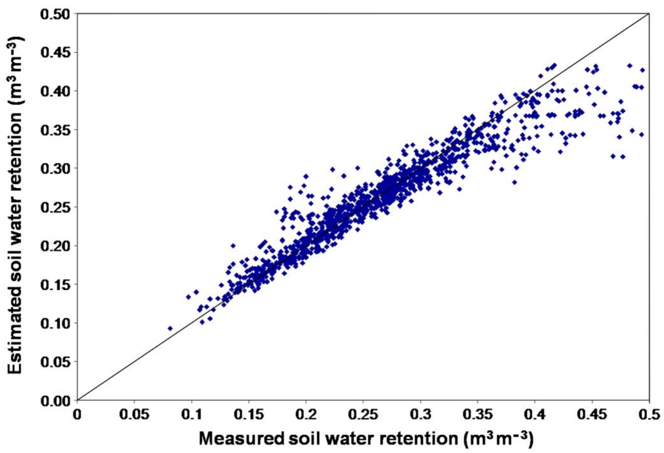

Fig. 3. Measured and estimated soil-water retention described by the Brooks-Corey function (poorest fitting)

a similar tendency. The BC (and VG) model (Fig. 3) underestimated retention in the same range. Underestimation was more prominent than overestimation, which was also reflected in statistical indices. The shrinking nature of black soils in the wet range probably was not adequately represented by the BC or VG parameters. Perhaps the sensitivity of the models in the wet range needs to be investigated. Key requirement for any parametric SWC expression should be parsimony (minmum parameters) to simplify parameter estimation and an accurate description of SWC behavior at the limits (wet and dry ends) while closely fitting the nonlinear shape of $\theta - h$ measurements. The advantages of using a given model lie in its complexity (number of parameters) and whether it needs to refit the experimental data or not (Webb et al. 2000). Irrespective of the model, deviations between measured and calculated moisture content were found only at lower suction ($< -1{,}000$ kPa) heads (dry range). In general, the wet range is relatively more important for computer modeling of field soils. It is in this region that most flow occurs. While comparing the four different soil-water retention functions, one can observe that all of them fitted the experimental data reasonably well. Obviously, each function adjusts its own set of parameters in order to minimize the error. The soil hydraulic functions of the VG model are comparatively difficult to calculate and do not lead to rapid numerical solution. It is also numerically expensive. The BC function can also be eliminated owing to its known limitations in fine-texture soils. Thus, the Campbell and CH functions could be a preferred choice for the study soils. It was concluded that performance of four functions (Campbell, CH, VG, and BC) should be analyzed further with more SWRC data. The samples were segregated according to USDA textural class; 111 samples were found to be of clay class, 34 belonged to silty clay, and rest (12) were of loam, sandy clay loam, and silty clay loam texture. The SWRC of the two major classes (clay and silty clay) were separated to identify the best-suited function. The statistical indices (Table 3) showed that the Campbell function was relatively better than three other functions in describing SWRC of clay samples. Patil et al. (2009) reported that the VG function was better suited to describe SWRC of seasonally impounded clay soils from the Jabalpur district, India (vertisols and intergrades).

Table 4. Statistical Indices to Judge Efficacy of Different Soil-Water Retention Functions in Describing SWRC Data of Black Soils (34 Silty Clay Samples)

Index/Pressure (-kPa)	33	100	300	500	800	1000	1500	Mean
BC								
RMSE	0.0325	0.0221	0.011	0.0114	0.0094	0.0112	0.0234	0.0173
d	0.87	0.94	0.98	0.98	0.98	0.97	0.88	0.94
ME	0.1361	0.0703	0.0249	0.0219	0.0225	0.0252	0.0521	0.0504
MAE	0.0175	0.0163	0.0091	0.0099	0.008	0.0082	0.0189	0.0126
Campbell								
RMSE	0.0384	0.0147	0.0121	0.0149	0.017	0.0136	0.0151	0.0180
d	0.95	0.99	0.99	0.98	0.97	0.98	0.97	0.97
ME	0.1331	0.0384	0.0388	0.0433	0.0484	0.0597	0.0374	0.0570
MAE	0.0267	0.0117	0.0095	0.012	0.0131	0.0094	0.0115	0.0134
CH								
RMSE	0.0236	0.0111	0.0109	0.0158	0.014	0.0129	0.0165	0.0150
d	0.96	0.99	0.98	0.96	0.96	0.96	0.91	0.96
ME	0.0756	0.0302	0.0294	0.044	0.033	0.0504	0.0395	0.0432
MAE	0.0154	0.0092	0.0088	0.012	0.0113	0.0076	0.0129	0.0110
VG								
RMSE	0.0308	0.015	0.0091	0.0105	0.0109	0.0097	0.0188	0.0150
d	0.89	0.98	0.99	0.98	0.98	0.98	0.91	0.95
ME	0.1322	0.0554	0.0205	0.0205	0.028	0.0321	0.0428	0.0474
MAE	0.0172	0.0109	0.0075	0.0086	0.0085	0.0071	0.0143	0.0106

J. Irrig. Drain Eng. 2012.138:177-184.

Table 5. Statistical Indices to Evaluate Performance of Hierarchical *k*-NN and Neural PTFs Developed

Input	RMSE	*d*	ME	MAE	R^2
ANN PTF					
SSC	0.0437	0.61	0.10	0.03	0.23
SSCBD	0.0401	0.58	0.10	0.03	0.34
SSCBDOM	0.0474	0.59	0.12	0.03	0.12
SSCOM	0.0395	0.6	0.10	0.02	0.22
k Nearest PTF					
SSC	0.0339	0.78	0.10	0.02	0.46
SSCBD	0.0426	0.71	0.09	0.03	0.38
SSCBDOM	0.045	0.67	0.10	0.03	0.29
SSCOM	0.040	0.77	0.11	0.06	0.45

Campbell and CH functions were also found to be well suited. The BC function was the worst performer. That study was confined to a smaller area and without any climatic heterogeneity. In this study, there were variations in soil genesis, climate, and topography. Thus, it could be argued that the Campbell function could describe SWRC of vertisols of varied origin and characteristics. As discussed earlier, it is the dry part of the retention curve that is difficult to describe well, irrespective of the function used. Therefore, preference for the Campbell model needs to be evaluated further because currently the findings are not fully convincing. Mixed results were obtained in silty clay soils (Table 4). The CH and VG functions had the lowest RMSE (0.0150), mean error was lowest in the CH functions, followed by VG. If the magnitude of RMSE and ME is given more importance, CH emerges as the most suitable function. It could be surmised that all four functions were within an acceptable range of RMSE, and hence a choice could be made by the user depending on the application or convenience.

Pedotransfer Functions

The performance of PTFs developed using *k*-NN and neural networks could be judged from the statistical indices (Table 5).

It could be observed that at the lowest input level (SSC), the performance of the *k*-NN PTF was relatively better (Fig. 4) as indicated by a lower an RMSE (0.0339) than an RMSE of 0.0437 at the same input level in a neural PTF. Other indices (*d*, ME, MAE, R^2) also confirmed the superiority of the *k*-NN PTF. The incremental addition of bulk-density data as an input variable did not improve performance of the *k*-NN PTF as evidenced by an increased RMSE (0.0426) in predicting AWC. The bulk density in vertisols is known to change with soil-water content. Perhaps, the data could not provide enough information on underlying relationship between SWR and BD, and hence the empirical relationship did not improve. The RMSE continued to increase with the addition of organic matter (Fig. 5) as an input variable with or without the addition of bulk density. However, the addition of organic matter alone as an additional input variable (Fig. 6) with textural composition exhibited a relatively lower RMSE (0.04). In general, *k*-NN PTFs had lower mean RMSE (0.0403) compared to neural PTFs (0.0426). Other statistical indicators also indicated that *k*-NN PTFs predicted AWC with greater accuracy irrespective of input/predictor variable level. Performance of neural PTFs exhibited improvement in RMSE (from 0.0437–0.0401) with inclusion of bulk density as a predictor variable. These results were expected because neural networks are known to show better predictive ability with an increase in the number of input variables. However, the highest RMSE (0.0474) was recorded at the maximum input level of

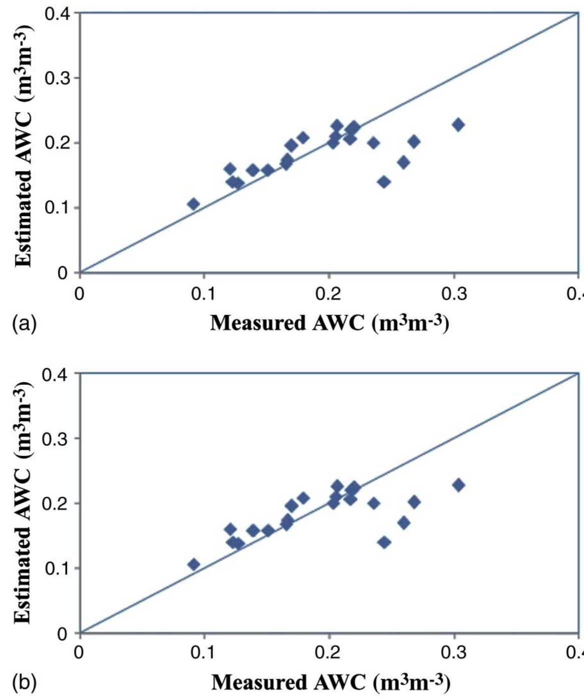

Fig. 4. (a) Measured and estimated available water capacity using *k*-NN PTFs with input of texture; (b) measured and estimated available water capacity using ANN PTFs with input of texture

texture, bulk density (BD), and organic matter (OM). The lowest RMSE and other indices at input SSCOM pointed to the importance of OM as a predictor variable. Evaluation of the *k* NN PTFs had also showed that the lowest input level (SSC) was adequate enough to get reasonable estimates of AWC. Thus, both techniques

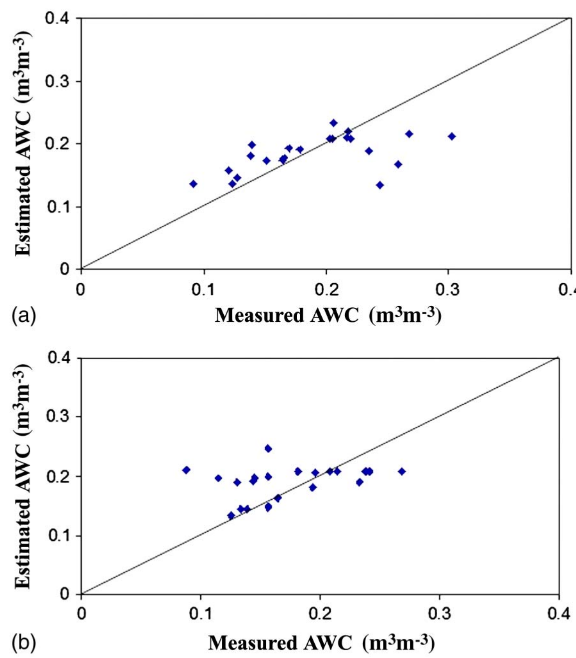

Fig. 5. (a) Measured and estimated available water capacity using *k*-NN PTFs with input of texture, bulk density, and organic matter; (b) measured and estimated available water capacity using ANN PTFs with input of texture, bulk density, and organic matter

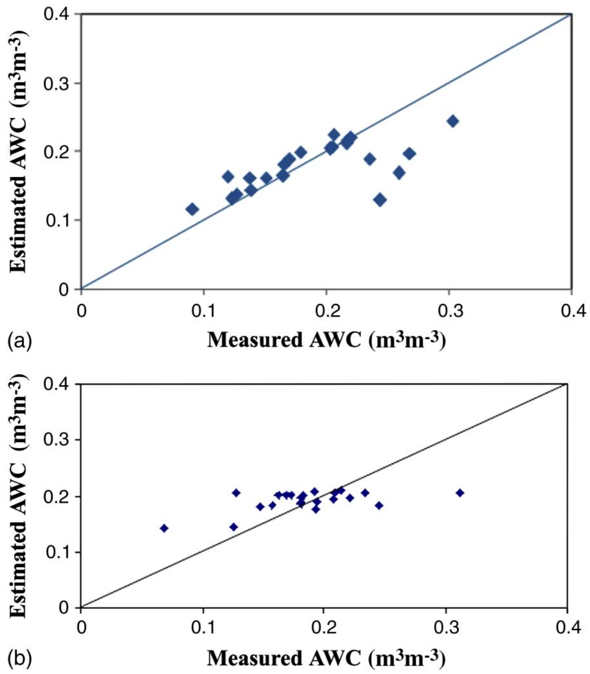

Fig. 6. (a) Measured and estimated available water capacity using k-NN PTFs with input of texture and organic matter; (b) measured and estimated available water capacity using ANN PTFs with input of texture and organic matter

underlined the importance of choosing a correct input variable for vertisols rather than the number of variables.

On comparison, it was evident that as a tool, k-NN performed better than neural networks with the additional advantage of simplicity in use and the possible appending of a development data set and hence could be a tool of choice. It is obvious that 26 profiles can provide only a glimpse of vertisols in India, and 77×10^6 ha is a vast area that needs representation by a far greater number of profiles. Precisely, due to this reason, the k-NN algorithm becomes important. As more and more SWRC data are acquired, the PTFs would need refinement for better predictions. Because ANN does not provide flexibility of appending data, PTFs would need to be redeveloped each time the data are added. On the other hand, k-NN PTFs could be refined without reprocessing. With proven acceptable accuracy, k-NN PTFs will be important for applications.

Conclusion

After evaluation, four of the 10 functions to describe SWRC of vertisols of India were recommended in order of preference: Campbell, Cass-Hutson, van Genuchten, and Brooks-Corey. However, the evaluation was based on limited data, and the recommended functions either overestimated or underestimated retention in the dry region of SWRC. It would be interesting to perform analysis with additional SWRC data before generalizing the recommendations. Neural regression and k-NN techniques of PTF development were evaluated. The best performing k-NN PTF to predict AWC required textural information as an input, while the best ANi matter in addition to the texture. Superior ability of k-NN PTFs in vertisols of India was noted. The study demonstrated that k-NN technique can be as competitive as widely used neural regression with the additional benefit of appending the development data as and when desired. The proposed PTFs could be useful in managing the vertisols of India.

Acknowledgments

Help rendered in preparation of this manuscript by Dr. Arun Chaturvedi and Shri. Pramod Tiwari is gratefully acknowledged. Their criticism helped in refining the manuscript to a great extent.

Notation

The following symbols are used in this paper:

d = degree of agreement; and
R^2 = coefficient of determination.

References

Ahmad, N., and Mermut, R. (1996). *Vertisols and technologies for their management*, Elsevier, Amsterdam.

Brooks, R. H., and Corey, A. T. (1966). “Porous media affecting fluid flow.” *J. Irrig. Div., Am. Soc. Civ. Eng.*, 92(2), 61–88.

Campbell, G. S. (1974). “A simple method for determining unsaturated conductivity from moisture retention data.” *Soil Sci.*, 117(6), 311–314.

Chang, W. Y., Liu, Z., Kang, L., Wang, Y., and Shi, M. (2004). “Parameters estimation of van Genuchten model for soil water retention curves using Matlab.” *Acta Pedol. Sin.*, 41(3), 380–386.

Driessen, P. M. (1986). “The water balance of the soil.” *Modelling of agricultural production: Weather, soils and crops. Simulation monographs*, Pudoc, Wageningen, Netherlands, 76–116.

Dudal, R. (1965). “Dark clay soils of tropical and sub-tropical regions.” *Agricultural Development Rep. 83*, Food and Agriculture Organization (FAO), Rome.

Farrel, D. A., and Larson, W. E. (1972). “Modeling the pore structure of porous media.” *Water Resour. Res.*, 8(3), 699–766.

Hutson, J. L., and Cass, A. (1987). “A retentivity function for use in soil water simulation model.” *J. Soil Sci.*, 38(1), 105–113.

Jackson, M. L. (1973). *Soil chemical analysis*, Prentice Hall India Pvt. Ltd., New Delhi.

Jain, S. K., Singh, V. P., and van Genuchten, M. T. (2004). “Analysis of soil water retention data using artificial neural networks.” *J. Hydrol. Eng.*, 9(5), 415–420.

Jongerius, A. and Heintzberger, G. (1975). *Soil micromorphology*, Methodology Soil Survey Institute, Wageningen, Netherlands.

Libardi, P. L., Reichardt, K., and Filho, V. F. (1979). “Analise da redistribuicao de agua visando a conductividade hidraulica do solo.” *Energ. Nucl. Agric.*, 1(2), 108–122.

Minasny, B., and McBratney, A. B. (2002). “The neuro-*m* method for fitting neural network parametric pedotransfer functions.” *Soil Sci. Soc. Am. J.*, 66(2), 352–361.

Minasny, B., McBratney, A. B., and Bristow, K. L. (1999). “Comparison of different approaches to the development of pedotransfer functions for water retention curves.” *Geoderma*, 93(3-4), 225–253.

Nemes, A., Rawls, W. J., and Pachepsky, Y. A. (2006a). “Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties.” *Soil Sci. Soc. Am. J.*, 70(2), 327–336.

Nemes, A., Rawls, W. J., Pachepsky, Ya. A., and van Genuchten, M. Th. (2006b). “Sensitivity analysis of the nonparametric nearest neighbor technique to estimate soil water retention.” *Vadose Zone J.*, 5(4), 1222–1235.

Nemes, A., Roberts, R. T., Rawls, W. J., Pachepsky, Ya. A., and van Genuchten, M. Th. (2008). “Software to estimate −33 and −1,500 kPa soil water retention using the non-parametric k-nearest neighbor technique.” *Environ. Modell Softw.*, 23(2), 254–255.

Patil, N. G., and Rajput, G. S. (2009). “Evaluation of water retention functions and computer program ‘Rosetta’ in predicting soil water characteristics of seasonally impounded shrink-swell soils.” *J. Irrig. Drain Eng.*, 135(3), 286–294.

Patil, N. G., Rajput, G. S., Nema, R. K., and Singh, R. B. (2009). "Predicting hydraulic properties of seasonally impounded soils." *J. Agric. Sci.*, 148(02), 159–170.

Schaap, M. G., Leij, F. L., and van Genuchten, M. T. (1998). "Neural network analysis for hierarchical prediction of soil hydraulic properties." *Soil Sci. Soc. Am. J.*, 62(4), 847–855.

Simmons, C. S., Nielsen, D. R., and Biggar, J. W. (1979). "Scaling of field measured soil water properties." *Hilgardia*, 47, 77–173.

van Genuchten, M. Th. (1980). "A closed form equation for predicting the hydraulic conductivity of unsaturated soils." *Soil Sci. Soc. Am. J.*, 44(5), 892–898.

Wagner, B., Tarnawski, V. R., Wessolek, G., and Plagge, R. (1998). "Suitability of models for the estimation of soil hydraulic parameters." *Geoderma*, 86(3–4), 229–239.

Webb, T. H., Claydon, J. J., and Harris, S. R. (2000). "Quantifying variability of soil physical properties within soil series to address modern land-use issues on the Canterbury Plains, New Zealand." *Aust. J. Soil Res.*, 38(6), 1115–29.

J. Irrig. Drain Eng. 2012.138:177-184.