Indian Journal of Agricultural Sciences **82** (1): 35–8, January 2012

Pedotransfer functions based on nearest neighbour and neural networks approach to estimate available water capacity of shrink-swell soils

NITIN GORAKH PATIL[1] and ARUN CHATURVEDI[2]

National Bureau of Soil Survey and Land Use Planning, Amravati Road , Shankaranagar, Nagpur, Maharashtra 440 033

Received: 16 August 2010; Revised accepted: 20 October 2011

ABSTRACT

Point pedotransfer functions (PTF) were developed by nearest neighbour (kNN) algorithm as an alternative to widely used artificial neural networks (ANN) for prediction of field capacity, permanent wilting point, and available water capacity of the seasonally impounded clay soils in central India using the available dataset. They were compared with parametric PTFs to estimate Van Genuchten parameters. Basic soil properties (texture, bulk density and organic carbon content) were related to the properties of interest. The root mean square error (RMSE) in predictions of the three properties (FC, PWP, AWC) by kNN PTFs ranged from 0.0237 to 0.0279 $m^3/m^3$with an average of 0.0265 m^3/m^3. The ANN PTFs performed relatively better with average RMSE 0.0223 m^3/m^3 and a range of 0.0141 to 0.0295 m^3/m^3. Magnitude of RMSE was relatively higher in VG parametric PTFs (0.032 m^3/m^3) followed by kNN (0.0237 m^3/m^3) and ANN (0.0223 m^3/m^3) PTFs. kNN algorithm provided viable alternative to neural regression with additional benefit of flexibility in appending reference database.

Key words: Available water capacity, Clay soils, K nearest neighbour, Neural networks, Pedotransfer functions, VG function

Pedotransfer functions (PTFs) have been used to obtain certain complex and expensive soil hydraulic parameters from other available or easily measurable soil properties in the last two decades (Jana and Mohanty 2010). Most of the PTFs reported in the literature pertain to estimation of soil water retention characteristics (SWRC). The PTF could be built to predict a point of interest on SWRC curve like field capacity (soil water retained at –33 kPa), permanent wilting point (soil water retained at –1500 kPa) etc. or PTFs to estimate parameters of a function describing the SWRC curve for continuous simulations. Modern PTFs are mostly derived using neural regression approach and a vast array of neural PTFs (e g. Jain *et al*. 2004; Patil *et al*. 2010) are found in the literature. Pattern recognition algorithms like k-Nearest Neighbor (k-NN) could replace neural regression resulting in PTFs that overcome the constraint faced by neural PTFs because reference database can be easily appended and the additional data can be used to improve accuracy of developed PTFs. Nemes *et al* (2006a, 2006b) have reported efficacy of kNN PTFs to predict soil water retention at –33 and –1500 kPa, commonly referred to as field capacity (FC) and permanent wilting point (PWP). They found that k-NN PTFs were as efficient as the PTFs developed using most advanced neural computing techniques. The objectives of this study were to*:* (i) establish point PTFs for the estimation of water retention points using neural regression and k nearest neighbour technique; (ii) establish parametric PTFs for the estimation of Van Genuchten (1980) water retention parameters, and (iii) compare estimates by different PTFs.

[1]Scientist (SG) (e mail: nitpat03@yahoo.co.uk), [2]Head (e mail: arunchat55@gmail.com), Division of LUP

MATERIAL AND METHODS

The study was conducted during 2004–09 at Nagpur. Soil properties data were collected and analyzed during first three years and subsequently soft computing was done. The data contain information on basic soil properties (Patil *et al.* 2009) and a nine point water retention characteristics of 41 soil profiles, (175 horizons of which 102 belonged to ‘clay’ texture as per USDA classification). The soils were classified as Vertisols and associated soils (Tomar *et al*. 1996) spread over nearly 5 mha. area. Point PTFs to predict FC/PWP/ AWC were developed using neural regression with the help of software ‘Neuropath’. For PTFs based on k nearest neighbour algorithm, software developed by Nemes *et al.* (2008) was used. Based on the earlier experience, (Patil *et al*. 2009), feed forward neural network model with three hidden nodes was preferred in ANN. Input information used in developing PTFs was as follows.

- 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 carbon content (SSCBDOC)

For deriving parametric PTFs, the data on SWRC were fitted to the van Genuchten (VG) model (Van Genuchten 1980) and relationships between basic soil data and the fitted parameters were established using neural regression.

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

Where, $\Theta(h)$ denotes the volumetric water content (L^3L^{-3}) at the corresponding soil-water matric head h (L), Θ_s saturated soil water content, Θ_r residual soil water content, α scaling parameter (/l), n a curve shape factor, and m an empirical constant that can be related to n. Performance of the developed PTFs was evaluated based on: (i) root mean square error (RMSE), (ii) index of agreement (d), (iii) maximum absolute error (ME), (iv) mean absolute error (MAE) and, (v) correlation coefficient (R^2). RMSE, d, ME, and MAE statistics were calculated using standard equations.

RESULTS AND DISCUSSION

Comparing kNN and ANN PTFs

It could be seen that the predictions by kNN PTFs improved with increased input as evidenced by decreasing RMSE (Table 1). The best predictions of FC (RMSE 0.0263 m^3/m^3) were obtained using the highest input level (SSCBDOC-fig.1a). ANN PTFS predicted FC with relatively greater accuracy (RMSE 0.0206 to 0.0258 m^3/m^3). Other indices also pointed to better performance by ANN PTFs. The best estimates were again obtained at the highest input level. The ANN PTFs were expected to improve with increase in input variables. The inclusion of information on bulk density and OC improved predictive ability of PTFs.

Evaluation of PTFs to predict PWP illustrated mixed trend as RMSE increased with hierarchical input of BD with SSC and then decreased with addition of OC (0.0249 to 0.0254 m^3/m^3) in kNN PTFs. However, the best estimates were obtained at highest input level (SSCBDOC-Fig 1b). Neural PTFs were relatively accurate (RMSE 0.0141 to 0.0210, d 0.75 to 0.77 ME 0.0365 to 0.0553 MAE 0.0137 to 0.017 m^3/m^3) in prediction of PWP. Again the ANN PTFs improved with increase in information level (more variables).

The best predictions were obtained with input of SSCBDOC. Thus, it could be inferred that FC/PWP of clay soils of the study area could be predicted using minimal information (textural composition) with acceptable accuracy. Similar trend of results was discernible in evaluation indices for PTFs to predict AWC with better performance by ANN PTFs, followed by kNN PTFs without great difference. Considering prediction of all the three properties, it was observed that the ANN PTFs performed relatively better with average RMSE 0.0223 m^3/m^3 and a range of 0.0141 to 0.0295 m^3/m^3. Other evaluation indices also corroborated the superior performance of ANN PTFs in predicting FC/PWP/AWC of clay soils. It was also apparent that the best

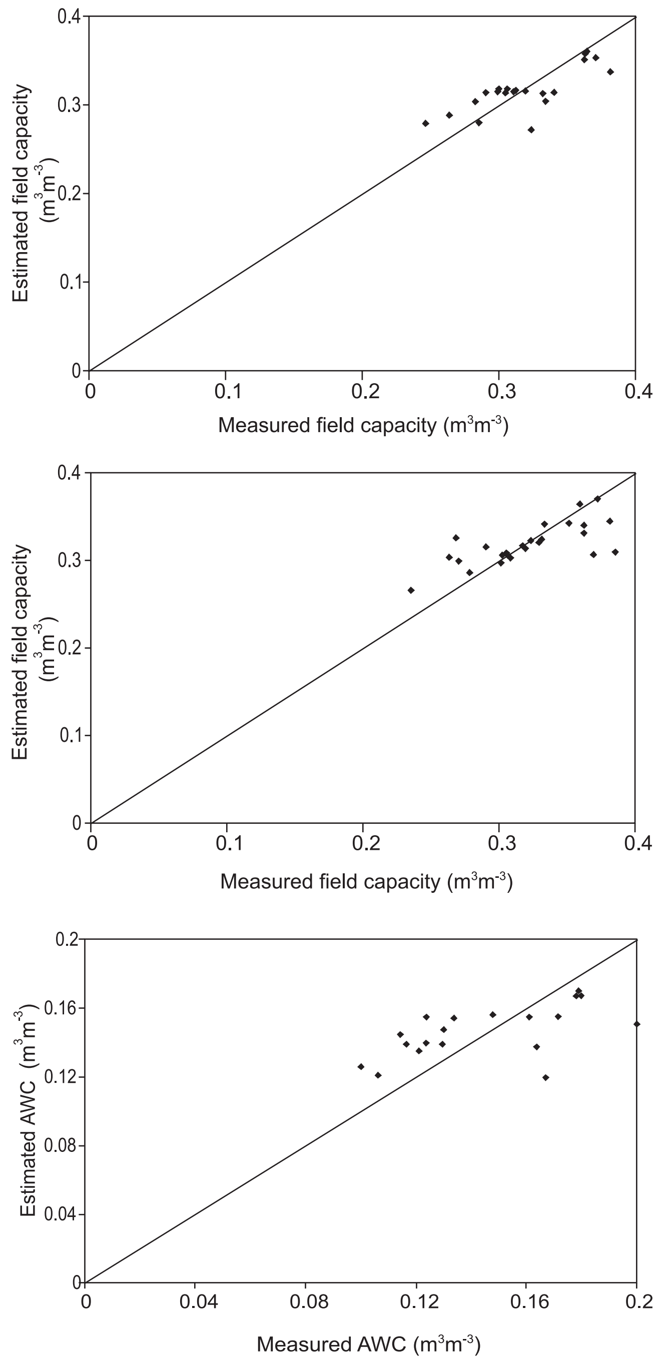

Fig.1 Measured and estimated field capacity using a) kNN PTFs b) ANN with input of texture, bulk density and organic carbon content and c) Measured and estimated AWC using parametric PTFs with the same input.

Table 1 Statistical evaluation indices to judge efficacy of PTFs

	Soil property	Input	RMSE (m^3/m^3)	d	ME (m^3/m^3)	MAE (m^3/m^3)
kNN PTFs	FC	SSC	0.0276	0.62	0.0611	0.0201
		SSCBD	0.0275	0.79	0.0673	0.0196
		SSCBDOC	0.0263	0.72	0.0597	0.0172
	PWP	SSC	0.0249	0.59	0.0704	0.0177
		SSCBD	0.0254	0.62	0.0782	0.0183
		SSCBDOC	0.0237	0.64	0.072	0.017
	AWC	SSC	0.0276	0.65	0.0717	0.0208
		SSCBD	0.0276	0.63	0.0699	0.022
		SSCBDOC	0.0279	0.65	0.0754	0.0206
	Mean		0.0265	0.66	0.0695	0.0193
	Max		0.0279	0.79	0.0782	0.022
	Min		0.0237	0.59	0.0597	0.017
ANN PTFs	FC	SSC	0.0258	0.84	0.0505	0.0182
		SSCBD	0.0246	0.84	0.0401	0.0211
		SSCBDOC	0.0206	0.76	0.0808	0.0196
	PWP	SSC	0.021	0.75	0.0365	0.0137
		SSCBD	0.0163	0.77	0.0466	0.0148
		SSCBDOC	0.0141	0.76	0.0553	0.017
	AWC	SSC	0.0224	0.76	0.0481	0.0176
		SSCBD	0.0263	0.73	0.0553	0.021
		SSCBDOC	0.0295	0.59	0.0603	0.0242
	Mean		0.0223	0.76	0.0526	0.0186
	Max		0.0295	0.84	0.0808	0.0242
	Min		0.0141	0.59	0.0365	0.0137
Parametric PTFs	FC	SSC	0.0279	0.63	0.0635	0.0239
		SSCBD	0.0311	0.49	0.0658	0.0265
		SSCBDOC	0.029	0.6	0.0657	0.0249
	PWP	SSC	0.0359	0.69	0.0941	0.0259
		SSCBD	0.0359	0.64	0.0948	0.0259
		SSCBDOC	0.0367	0.66	0.0933	0.0272
	AWC	SSC	0.0336	0.5	0.08	0.0269
		SSCBD	0.0339	0.47	0.0817	0.0284
		SSCBDOC	0.0349	0.49	0.0851	0.0279
	Mean		0.0332	0.57	0.0804	0.0264
	Max		0.0367	0.69	0.0948	0.0284
	Min		0.0279	0.47	0.0635	0.0239

FC-Field capacity; PWP, permanent wilting point; AWC, available water capacity

performing PTFs needed textural information (SSCBDOC) as an input. Thus data on bulk density or organic carbon content was desirable for obtaining good estimates.

Point and parametric PTFs

The predictions by point PTFs were relatively accurate than the predictions obtained using parametric PTFs. The best predictions by parametric PTFs were obtained using minimal input level (Table 1, Fig 1c) of textural information related to VG parameters through neural regression. The VG parametric PTFs exhibited marginally higher RMSE (0.0332 m^3/m^3) than either of the kNN or ANN PTFs. Interestingly, one of the popular generic PTF Rosetta (Schaap *et al*. 2001) was reported to predict AWC of these soils with RMSE 0.0332 and 0.0331 m^3/m^3 with input of SSC and SSCBD respectively (Patil *et al.* 2010). Magnitude of mean error was relatively higher in VG parametric PTFS (0.08 m^3/m^3) followed by kNN (0.069 m^3/m^3) and ANN (0.052 m^3/m^3) PTFs. All the indices confirmed that point PTFs were a preferred choice over parametric PTFs. Borgesen and Schaap (2005) have reported similar findings in their study on Danish soils. Sharma *et al.* (2006) have reasoned that improved prediction for one of the VG parameters does not necessarily mean better performance in predicting the water contents at different matric potentials due to nonlinear nature of the hydraulic functions. It could be inferred that for predicting FC and PWP, relative preference could be accorded to point PTFs. Since, estimates by Rosetta were also acceptable, there were four options (i) easy to use kNN PTFs, (ii) ANN PTFs requiring relatively greater efforts but better predictive ability, (iii) parametric PTFs, and (iv) generic PTF Rosetta , later two with almost identical accuracy and applicability in simulations requiring continuous relationship between pressure and soil water retention.

The best ANN as well as kNN PTFs in general required maximum input information (SSCBDOC) as against the lowest information requirement in parametric PTFs. Several studies have used information like bulk density as an input to improve the accuracy of PTFs (eg, Rawls *et al.* 1992; Williams *et al.* 1992). Our findings are in agreement with these reports. Parametric (VG) equation could not precisely match observed water contents; PTFs based on such parameters resulted in higher errors than point PTFs. The error could partly be attributed to shrink swell nature of soil and their retention behaviour in dry region where PWP may be reached much earlier than the –1 500 kPa suction pressure.

The study attempted comparative evaluation of point and parametric PTFs to predict three soil hydraulic properties namely FC/PWP/AWC of clay soils. The point PTFs competed well with parametric PTFs implying that measurement of soil water retention characteristics (multi-point data) were not essential at least to build PTFs to predict the three soil hydraulic properties. Comparison of kNN and ANN technique showed that ANN was relatively better in terms of minimizing prediction error but kNN was equally competitive with additional advantage of simplicity in use and possible appending of development dataset and hence could be a tool of choice to predict FC/PWP/AWC of the study soils.

REFERENCES

Borgesen C D and Schaap M. 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. *Geoderma* (**127**): 1-2: 154–67.

Jana R B and Mohanty B. 2010 Enhancing PTFs with remotely sensed data for multi-scale soil water retention estimation. *Journal of Hydrology* **399** (3-4): 201–11.

Jain S K, Singh V P and Van Genuchten M. Th. 2004. Analysis of soil water retention data using artificial neural networks. *Journal of Hydrologic Engineering* **9:** 415–20.

Nemes A, Roberts R T, Rawls W J, Pachepsky Ya A and van Genuchten M Th. 2008. Software to estimate -33 and -1500 kPa soil water retention using the non-parametric *k*-Nearest Neighbor technique. *Environmental Modeliling and Software* **23**(2): 254–5

Nemes, Rawls W J, Pachepsky Ya A and van Genuchten M Th. 2006a. Sensitivity Analysis of the Nonparametric Nearest Neighbor Technique to Estimate Soil Water Retention *Vadose Zone Journal* **5**: 1222–35.

Nemes, Rawls W J and Pachepsky Y A. 2006b. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Science Society of America Journal* **70**: 327–36.

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. *Journal of Irrigation and Drainage Engineering* **135** (3): 286–94.

Patil N G, Rajput G S, Nema R K and Singh R B. 2010. Calibration and evaluation of pedotransfer functions to estimate available water capacity of seasonally impounded shrink-swell soils of Central India. *Archives of Agronomy and Soil Science* **148** : 159–70.

Rawls W J, Ahuja L R and Brakensiek D L. 1992. Estimating soil hydraulic properties from soils data. (*in*) *Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils. Proceedings of International Workshop on Riverside, CA,* pp. 329–40. M.Th. van Genuchten *et al.* (Eds). 11–13 October, 1989, University of California, Riverside.

Sharma Sanjay K, Mohanty Binayak P and Jianting Zhu. 2006. Including topography and vegetation attributes for developing pedotransfer functions. *Soil Science Society of America Journal* **70**:c 1430–40.

Schaap M G, Feike J L, Martinus L, and Van Genuchten M Th. 2001. Rosetta a computer program for estimating soil hydraulic parameters with hierarchial pedotransfer functions. *Journal of Hydrology* **251**: 163–76.

Tomar S S, Tembe G P, Shrma S K, Bhadauria U P S and Tomar V S. 1996. Improvement of physical conditions of black soils in Madhya Pradesh J N K K V, Jabalpur.

Van Genuchten M Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science. Society of America. Journal* **43**: 892–8.

Williams R D, Ahuja L R and Naney J W. 1992. Comparison of methods to estimate soil water characteristics from texture, bulk density and limited data, *Soil Science* **153**. 172–84.