Application of neural network pedotransfer functions to calculate soil water retention curve

A A Belik¹², A G Bolotov³, E V Shein¹, A A Kokoreva¹², A A Levin⁴ and V Yu Patrushev⁴

¹ Faculty of Soil Science, Lomonosov Moscow State University, GSP 1, Leninskie Gory, Moscow 119991, Russian Federation
² Russian Scientific-Research Institute of Phytopathology Bolshiye Vyazemy, Moscow region, Russian Federation
³ Meteorology and Climatology Department, Russian State Agrarian University - Moscow Timiryazev Agricultural Academy, 127550, Moscow, Timiryazevskaya st, 49, Russian Federation
⁴ Faculty of Environmental Engineering, Altai State Agrarian University, 656031, Barnaul, Merzlikina st, 8, Russian Federation

*E-mail: belikalexandra@gmail.com

Abstract. The results of soil hydrophysical parameters modeling by neural networks are considered. Modern computing facilities using pedotransfer functions provide simplicity, speed and cheapening of obtaining soil characteristics by mathematical calculations. The initial data are the basic properties of the soil, which are stably determined. The analysis of errors in the calculation of hydrophysical parameter was conducted. Obtained neural networks most accurately reproduce hydrophysical parameters.

1. Introduction

Hydrophysical analysis often includes an assessment of water infiltration into the soil, water transfer and retention by soil matrix. To use the mathematical models of substances migration it is necessary to know the soil hydrophysical parameters. Obtain the values of moisture pressure and hydraulic conductivity in field or laboratory conditions is difficult, expensive, and often impractical in soil hydrological research. The high correlation between soil properties and hydrophysical parameters can promote a fairly accurate assessment of many analyzes of biospheric phenomena and environmental management. For example, to study the problems of subsoil compaction [1]. Therefore, the computational methods based on these parameters’ dependence on the soil basic physical properties are a good alternative for the experimentally obtained hydrophysical parameters.

Functional dependences allowing converting information about the soil (pedo) properties into the information about its transport characteristics are called pedotransfer functions (PTF) [2]. The modern computational base using PTF provides simplicity, speed and cheapening obtaining soil characteristics through mathematical calculations, since the initial data are the soils basic properties, which are stably determined. Thus, a pedotransfer model tested for different soil classes can be provided by calculating the necessary characteristics.

2. Objects and methods
To study the water retention curve the centrifugation method in the pressure range of 2.5-10710 Pa was used [3]. The experimental obtained water retention curves for the main diagnostic horizons of the Altai Region zonal soils were approximated by the function of the water retention curve model van Genuchten in the RETC software [4]. The value of RMSE was 2-16%. The following hydrophysical parameters were obtained: $\theta_r$ - the residual water content [cm$^3$/cm$^3$]; $\theta_s$ - the saturated water content [cm$^3$/cm$^3$]; $\alpha$ - the inverse proportion to the bubbling pressure [1/Pa]; $n$ - the index of pore size distribution characterizing the slope of the water retention curve [-]; $K_r$ - saturated hydraulic conductivity [cm/day]. To obtain $K_r$, simultaneously with the measurement of soil moisture pressure, the moisture conductivity functions were determined from the kinetics of centrifugation. The dependence of the residual moisture of the sample on time at a constant speed of rotation of the sample is approximated by a parabolic function using a linear relaxation model [5]. The PTF as a characteristic of the Altai Region soils was obtained. The data set consisted of 810 experimental water retention and hydraulic conductivity curves. And the same time curves with simultaneous data on the particle size distribution (PSD), bulk density and soil organic matter (SOM) content were obtained. Soil PSD was analyzed by pipette-method after dispersion with sodium-pyrophosphate. Soil density was determined by drilling method. SOM was determined by an AN-7529M express analyzer [6].

The following hydrophysical parameters and physical properties are described in the articles [7-9]. PTFs were based on training array, which is 2/3 of the total sample size. The accuracy of the forecast was estimated according the test sample, which is 1/3 of the total sample size.

The authors used a functional-parametric method in which regression dependencies relate to the parameters of the water retention curve approximation to the basic soil properties [6]. The predictors were the bulk density, the particle size distribution, the SOM content, and the van Genuchten parameters of the water retention model were used as the response variable. PTFs were obtained using the artificial neural network (ANN) method created based on the Neural Networks Statistica software. Before creating a neural network, an array of related data $\theta_r$, $\theta_s$, $\alpha$, $n$ with thermal and hydrophysical characteristics, as well as with the basic physical properties of the soil, was divided into a training set used for network training and a control set for checking the quality of the network operation. This set was not used in the learning process. For verification, a cross-validation was performed (verification in the created network with data from the control set - test sample). The procedure of generalization with the control error monitoring was performed, excluding the “over-fitting” and re-training in the absence of cross-checking (for several data arrays).

In soil physics, to express complex dependencies between hydrophysical and basic soil properties, the most commonly used type of ANN is a multilayer perceptron [10-11]. The advantages of ANNs compared with regression approaches are in their ability to approximate almost any continuous function (if there is a sufficient number of levels, neurons, and interconnections), so the researcher does not need to accept any hypotheses about the model in advance. In addition, ANNs cope with the “curse of dimension”, which does not allow modeling linear dependencies for many variables [6]. At the network creation stage, regression of the task type was chosen while continuous functions with line-by-line removal of missing values was chosen as input and output elements’ type, since neural network is sensitive to data passes.

When choosing the network type, a combination of linear, nonlinear and radial based functions (RBF) and the three-layer perceptron with further selection of the best network was used.

The complexity of the network was determined by the number of hidden elements: in the radial-basis function from 1 to 138; the number of layers (from 1 to 10) in a multilayer perceptron (MP) simulating the response function using “sigmoid slopes” functions. The analysis duration was limited to 10 networks. The criteron of selecting one network from the entire set of the created networks was determined by the smallest error in the control sample, which did not exceed 12%, and in most cases was not more than 5%. During the work, the comparison of several networks demonstrated that as the number of elements on the hidden network layer increases, the accuracy of the results grows as well.
To predict $\theta_r$, the network has a direct propagation structure with 5 input variables, 6 elements on the hidden layer and one element on the output layer. The input parameters were: the particle size distribution, the bulk density and the SOM content. The created networks successfully recognized the structure of the training set and are suitable to be used in predicting the values of van Genuchten parameters. The PTF is stored in a file format C programming language, which is ready for integration with other applications.

3. Research results
The neural network underestimates the calculated values of $\theta$, compared with those observed for sandy clay and clay soils, which is proved by the angle the approximation line of inclination under 45 degrees (figure 1). The data analysis showed that the $\theta$ calculation of sandy soils using a neural network is more accurate than the calculation based on regression ratios [12]. One of the methods to increase the models’ accuracy is removing data about organic matter content in the underlying horizons out from the experimental dataset. For neural network analysis, when deleting data from SOM content < 0.6% of the sample, the calculation error decreased from 23% to 21%.

“Teaching” ANN through soil samples of a single climatic zone (a colossal forest-steppe) has improved the forecast accuracy, $S = 18\%$ (figure 2). The average value of the mean error is 0.0092, which allows to conclude that systematic component of the error is absented.

![Figure 1. Correlation field of calculated and experimental values for $\theta_r$ parameter, $S = 23\%$. The neural network MP 8:8-11-7-1:1.](image1)

![Figure 2. Correlation field of calculated and experimental values of $\theta_s$ parameter, $S = 18\%$. The neural network MP 6:6-11-7-1:1.](image2)

Using neural network analysis to calculate the value of $\theta_s$, revealed that the most accurate model was obtained through a linear neural network with two inputs (figure 3). The training error as well as controlling and testing errors of the constructed models varied in the range of 5.3-7.0%; 5.6-7.2%; 5.6-6.4% respectively. Wherein, the lowest error values were shown by the linear ANN. The neural networks’ architecture ($\theta_s$) is much simpler than the ones of the ANN ($\theta_r$), so the number of inputs is about 2-3 instead of 7-8, and the number of hidden neurons is 6-9 instead of 7-11.

Out of many neural networks obtained during the test, it was the multilayer perceptron of the MP architecture 4:4-6-1:1 and the MP 5:5-7-1:1 which reproduced the most accurate parameter $\alpha$. The neural network with such input parameters as physical clay, SOM and bulk density composition reproduced the parameter $\alpha$ with the lowest accuracy so it is not given here. The neural network with such input parameters as sand content, coarse silt, physical clay, SOM content and the bulk density reproduced the parameter $\alpha$ with the highest accuracy (figure 4).
Increasing the inputs up to 5 (adding a sand fraction) increased the prediction accuracy of $\alpha$ in comparison with the four-input multi-layer perceptron. The parameter $n$ was modeled using neural networks based on the following input variables: the sand fraction content, physical clay, SOM content, and the bulk density (figure 5).

The lowest accuracy in the $K_s$ calculation was obtained using a neural network based on a radial basis function with 3 inputs and 16 hidden neurons. The neural networks input parameters were the content of the sand fraction, physical clay and the bulk density (figure 6).

4. Conclusion
Summarizing: checking the zero mean property showed it to be close to zero in all the considered cases, therefore the obtained PTFs do not contain a constant systematic error and are adequate by the zero mean criterion. The obtained PTFs can be used to forecast soil hydrophysical parameters.

Acknowledgment
This work was supported by the Russian Scientific Foundation, project no. 16-16-04014 and by RFBR according to the research projects no.18-44-220007, 19-29-05112.
References
[1] Simsek U, Shein E V, Mikailsoy F, Bolotov A G and Erdel E 2019 Subsoil Compaction: The Intensity of Manifestation in Silty Clayey Calcic Pantofluvic Fluvisols of the İğdır Region (Eastern Turkey) Eurasian J. Soil Sci. **52** 296-99
[2] Bouma J A J and van Lanen 1987 Transfer functions and the shold values: From soil characteristics to land qualities (The Netherlands: ITC Publ., Enschede)
[3] Smagin A V, Sadovnikova N B and Ali M M B 1998 The determination of the primary hydrophysical function of soil by the centrifuge method Eurasian Soil Sci. **31** 1237-44
[4] van Genuchten M Th, Leij F J and Yates S R 1991 The RETC code for quantifying the hydraulic functions of unsaturated soils (Riverside, CA: US Salinity Laboratory) p 216.
[5] Smagin A V and Sadovnikova N B 1994 Impact of Strongly Swelling Hydrogels on Water Holding Capacity of Light-Textured Soils Eurasian Soil Sci. **27** 26-34
[6] Shein E V et al 2007 Theories and Methods of Soil Physics (Moscow: Grif and K) p 616 (in Russian)
[7] Bolotova A G, Shein E V and Makarychev S V 2019 Water retention capacity of soils in the Altai Region Eurasian Soil Sci. **52** 187-192
[8] Makarychev S V and Bolotov A G 2017 Conductive and steam-diffuse constituents of thermotransfer in different soil moisture contents: case study of the Altai Region’s soils Eurasian J. Soil Sci. **6** 44-50
[9] Makarychev S V and Bolotov A G 2016 Structural-functional concept of thermophysical condition of the soils of Altai Region Eurasian J. Soil Sci. **4** 279-284
[10] Pachepsky Ya A, Timlin D and Varallyay G 1996 Artificial neural networks to estimate soil water retention from easily measurable data Soil Sci. Soc. Am. J. **60** 727-733
[11] Schaap M G, Leij F L and van Genuchten M Th 1998 Neural network analysis for hierarchical prediction of soil hydraulic properties Soil Sci. Soc. Am. J. **62** 847-855.
[12] Ilinich V V, Bolotov A G, Shein E V and Makarychev S V 2018 Assessment of surface moisture in the catchment area on the base of modelling the hydrological properties of soils EPiC Ser. Eng. **3** 931-935