Evaluation of Soil-Parameters with Diffusion-Neural-Network

Zoltán Makó, Szilárd Máté

Sapientia University
530104 Miercurea, Ciuc, P-ta Libertăţii 1, Romania
e-mail: makozoltan@sapientia.siculorum.ro, mateszilard@sapientia.siculorum.ro

Abstract: Pedotransfer (PT) function can be defined as a predictive function of certain soil properties from other easily-, routinely-, or cheaply measured properties. The usual step in deriving PT functions is by forming empirical relationships between basic soil properties and parameters to be predicted.
An Artificial Neural Network is a learning machine whose function depends on the training samples. According to the principle of information diffusion we can increase the certainty of the determined relation if we multiply the number of the training samples with the help of an appropriate information scattering function. Neural networks trained in this manner are called diffusion-neural-networks (DNN).
In this paper we present a DNN, which can be trained to calculate the pH value of the soil using samples from the Ciuc-basin. With the weights of the trained network we construct a pH-value calculating PT function.

Keywords: pedotransfer function, diffusion-neural-network, soil-evaluation

1 Introduction

Pedotransfer functions have important role in building models for examining soil properties. The term pedotransfer function (PTF) was coined by Johan Bouma (in 1989) as translating data we have into what we need.

In soil science this means a function which can be defined as predictive functions of certain soil properties from other more available, easily, routinely, or cheaply measured properties (McBratney et al, 2002).

The usual step in deriving PT functions is by forming empirical relationships between basic soil properties and parameters to be predicted. This can be achieved by various mathematical methods, such as multiple linear regressions (Wösten et al, 1995). A recent approach for fitting PT functions is to use artificial neural networks (Pachepsky et al, 1996; Schaap et al, 1998).

In Romania, soil-evaluation is made in virtue of the 125/1999 Ministry Ordinance. The methodology, the calculating models and the soil-evaluation was elaborated in 1987 by the Soil and Agrochemistry Institution. The soil value points are calculated from 17 ecological parameters (average annual temperature, average annual precipitation fall, gleyzation, surface-water gleyzation, particle-size distribution, pollution, average slope, erosion, ground water level, flood risk, pore space, carbonate capacity, pH, physiologically useful volume, humus capacity and moisture excess) with the next formula:

$$S_v = (k_1 \times k_2 \times ... \times k_{17}) \times 100 \qquad (1)$$

where the k_i parameters depends from the subjectivity of specialists and they are selected very circuitously from many page wide tables.

The base of calculation in this Ordinance is the country's first 8 yield of out-turns resulting a soil value point in the 0-100 interval. For this in intermountain, cold and humid basins like the Ciuc-basin due to climate prohibitive factors soil value points are below 50, resulting only III, IV and V class soil.

The pH value is a determining parameter for the soil-value. In this paper we give a method for estimating pH values in unknown points based on a few sample points. The study area is a 6x10 km zone in the Ciuc-basin (Figure 1).
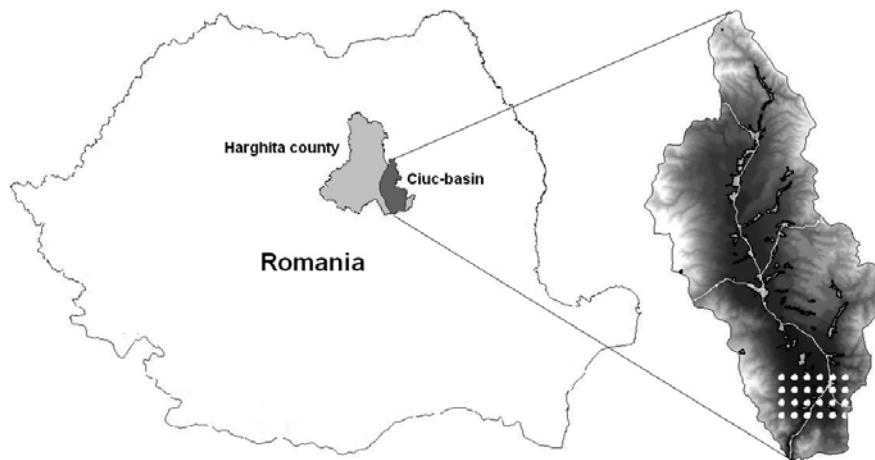

Figure 1
Location of the sample points

For the prediction of the values we use the principle of information diffusion. In the Section 4 of the paper we present a Difussion-Neural-Network (DNN) which after a learning process is capable to estimate the pH value with 0.95 probability.

2 The Diffusion of Information

Let $A = \{\mathbf{x}_k \in X \ / \ k = 1,2,...,n\}$ be a sample of data in a given normed space X that comes from the observation of some phenomenon. Let us denote a real relation with R. The method that defines the R from sample A is called an operator. Examples for operators: data series analysis, correlation examination, hypothesis examination, least squares method, the method of artificial neural networks, etc.

Definition 2.1 Let R be a relation in X. The sample A is a *correct-data set* to R on universe $U \subseteq X$ if there exists an operator γ such that we can obtain a relation $R(\gamma,A)$ equal to the restriction of R at U.

Definition 2.2 Let R be a relation in X. The sample A is an *incomplete-data set* to R on universe $U \subseteq X$ if there doesn't exist an operator such that we can obtain the restriction of R at U from A.

Definition 2.3 We consider a division U_j, $j = 1,...,m$ of universe U, i.e.

$$U = \bigcup_{j=1}^{m} U_j, \ U_j \cap U_k = \varnothing \ \text{if} \ j \neq k.$$

The *characteristic function of the division* U_j is $\chi_m : A \times U \rightarrow \{0,1\}$, where $\chi_A(\mathbf{x}_k, \mathbf{u}) = 1$ if $\mathbf{x}_k \in U_j$ and $\chi_A(\mathbf{x}_k, \mathbf{u}) = 0$ if $\mathbf{x}_k \notin U_j$, for all $\mathbf{u} \in U_j$.

The characteristic function is replaceable with membership function $\mu : A \times U \rightarrow [0,1]$. In this case, the value $\mu(\mathbf{x}_k, \mathbf{u})$ shows how far the sample's element $\mathbf{x}_k$ is in set U_j.

Definition 2.4 The family of membership functions $\mu(:, \mathbf{u}_j) : U \rightarrow [0,1], j = 1,...,m$ is a *fuzzy division* of the U.

Since, the membership functions $\mu(:, \mathbf{u}_j)$ diffuse the information among the fuzzy sets U_j, hence the relations searching methods (operators) that use these membership functions will be called *information-diffusion methods* (operators).

Definition 2.5 Let $A = \{\mathbf{x}_k \in X \ / \ k = 1,2,...,n\}$ be a sample of universe U. The function $\mu : A \times U \rightarrow [0,1]$ is a *scattering function of the information*, if

    i) $\mu(\mathbf{x}_k, \mathbf{x}_k) = 1$, for all $\mathbf{x}_k \in A \cap U$;

    ii) for all $\mathbf{x}_k \in A$ and for all $\mathbf{u}, \mathbf{v} \in U$, if $\|\mathbf{x}_k - \mathbf{u}\| \leq \|\mathbf{x}_k - \mathbf{v}\|$ then

        $\mu(\mathbf{x}_k, \mathbf{u}) \geq \mu(\mathbf{x}_k, \mathbf{v})$.

For all elements $\mathbf{x}_k$ of the sample A the scattering function define a fuzzy number with centre in $\mathbf{x}_k$ and membership function $\mu(\mathbf{x}_k, :) : U \rightarrow [0,1]$. The simplest scattering function is $\mu_t = \chi$. This function will be called *trivial scattering function* of the information.

The scattering function of the information shows how far the data **u** can be the correct-data of a phenomenon. For example, if **u** is in sample A then **u** is totally correct-data of the phenomenon. Using the scattering function μ the sample A can be expand with new elements and so we get a sample notated by $A(\mu, U)$ with elements $(\mathbf{x}_k, \mathbf{u}_j, \mu(\mathbf{x}_k, \mathbf{u}_j)) \in A \times U \times [0,1]$, where $\mathbf{u}_j \in U$, $j = 1,...,p$.

If $X = \mathbb{R}^n$ then it possible to define scattering function with help of quasi-triangular fuzzy numbers (Kovács, 1992). Let $p \in [1, +\infty]$ and $g: [0, 1] \to [0, +\infty]$ be a continuous, strictly decreasing function with boundary properties $g(1)=0$ and $\lim_{t \to 0} g(t) = g_0 \leq +\infty$. The triangular norm generated by g^p is given by the formula $T_{gp}(x,y) = g^{[-1]}\left(\left[g^p(x) + g^p(y)\right]^{1/p}\right)$, where

$$g^{[-1]}(x) = \begin{cases} g^{-1}(x) & \text{if } 0 \leq x \leq g_0, \\ 0 & \text{if } x > g_0. \end{cases}$$

We fuzzyfied all elements of sample A, i.e. for all components x_{ki} of vector $\mathbf{x}_k \in A$ we assign a quasi-triangular fuzzy number $<x_{ki}, \lambda(x_{ki})>$ with spread $\lambda(x_{ki}) > 0$, $i = 1,...,n$.

As follows from the definition of T_{gp}-Cartesian product the scattering function of information is given by

$$\mu\left((x_{k1}, x_{k2},...,x_{kn}),(u_1, u_2,...,u_n)\right) =$$

$$\begin{cases} g^{-1}\left(\left[\left[\left(\frac{|x_{k1} - u_1|}{\lambda(x_{k1})}\right)^p + ... + \left(\frac{|x_{kn} - u_n|}{\lambda(x_{kn})}\right)^p\right]^{1/p}\right) & \text{if} \quad |u_i - x_{ki}| \leq \lambda(x_{ki})g_0, i = \overline{1,n}, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Let R be a relation on universe $U \subseteq X$ and γ be an operator. If we are using the sample $A = \{\mathbf{x}_k \in X / k = 1,2,...,n\}$ to estimate the relation R then our method is a *nondiffusion estimator*, and if we are using the sample $A(\mu, U)$, where μ is a nontrivial information scattering function, then our method is a *diffusion estimator*. The trivial information scattering function yield a nondiffusion estimator

Theorem 2.1 (Principle of information diffusion - Huang and Shi, 2002) Let R be a relation on universe $U \subseteq X = \mathbb{R}^n$, where U is a convex set. Let $A = \{\mathbf{x}_k \in X / k = 1,2,...,n\}$ be a deterministic sample for estimation of R on universe $U \subseteq X$. We assume that γ is the best operator of relation R for some measurement of the error. The sample A is incomplete-data set of the relation R on U if and only if there exists a nontrivial information scattering function μ such that if we apply the operator γ to fuzzified sample $A(\mu, U)$, then we get a better estimation of R.

The proof of this theorem it is found in (Mako, 2006).

3 The Approximation Property of BP Artificial Neural Network

The neural network can be understood as a mapping $f : R^n \to R^m$, defined by $\mathbf{y} = f(\mathbf{x}) = g(W\cdot\mathbf{x})$, where $\mathbf{x}$ is the input vector, $\mathbf{y}$ is the output vector, W is the weight matrix and g is the activation function. The mapping f can be decomposed into a chaining of mappings; the result is a multi-layer network $R^n \to R^p \to R^q \to \ldots \to R^m$. The algorithm for computing W is often called the training algorithm. The most popular neural network are the multi-layer back-propagation networks whose training algorithm is the well-known gradient descendent method. Such networks are called back-propagation (BP) networks.

An artificial neural network is a learning machine whose function depends on the training examples. So, the machine does not recognize the real relation but it determines a numerical relation among the state parameters.

A number of authors have discussed the universal approximation property of BP networks. For example, in 1989 G. Cybenko showed that any $f: [a,b] \to R$ continuous function can be approximated by a neural network with one internal hidden layer using sigmoidal activation function (Cybenko, 1989). Also in 1989 K. Hornik et al. proved that the multi-layer networks can approximate the continous function to any degree of accuracy, i.e. multi-layer networks have the universal approximation property (Hornik et al., 1989). After that, in 1995 J. Wray and G. G. R. Green showed that, the universal approximation property does not hold in practice for networks implemented on computers (Wray and Green, 1995).

According to the principle of information diffusion we can increase the certainty of the determined relation if we multiply the number of the training examples with the help of an appropriate information scattering function. Neural networks trained in this manner are called *diffusion-neural-networks* (DNN - Huang and Shi, 2002; Huang and Moraga, 2004).

4 The Approximation of Soil Parameters with DNN

As we mentioned above, calculating the soil-value is based upon knowing the exact values of the determining parameters. In default of these parameters the soil-value can't be calculated with acceptable accuracy.

In Figure 2 we can see that increasing the number of sample points leads to contour lines with significantly different shape.
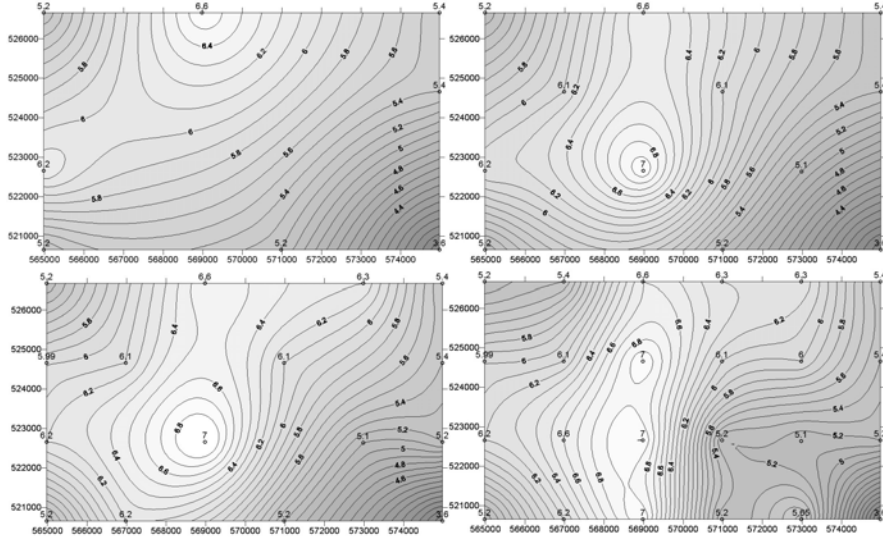
Figure 2
Contour lines of pH obtained from 8, 12, 16 and 24 sample points

The few sample points raises the question that how near are to reality the results obtained from 24 sample points? Whether appears or not new spots if we increase the number of sample points?

According to the principle of information diffusion we can increase the certainty of the pH value if we multiply the number of sample points with the help of an appropriate information scattering function.

In interest of this we proceed in the following way. Assuming that the pH value has normal distribution, we generate the information diffusion with the $g: [0, 1] \rightarrow [0, +\infty]$, $g(x) = \sqrt{\ln(1/x)}$ function and with the $p = 2$ number. In this case the information diffusion function is

$$\mu\left((x_k, y_k, pH_k),(x, y, pH)\right) = \exp\left(-\left(\frac{|x - x_k|}{\lambda(x_k)}\right)^2 - \left(\frac{|y - y_k|}{\lambda(y_k)}\right)^2 - \left(\frac{|pH - pH_k|}{\lambda(pH_k)}\right)^2\right)$$

Let $A = \{(x_k, y_k\ pH_k) \mid k=1,2,...,14\}$ be a given sample. The sample set is choosen in order to contain the minimum and the maximum pH value and the peaks value of the region (Table 1).

The X coordinate	The Y coordinate	Ph value
520657.00	564977.00	5.20
520657.00	566977.00	6.20
520657.00	568977.00	7.00
520657.00	570977.00	5.20
520657.00	574977.00	3.60
522657.00	564977.00	6.20
522657.00	568977.00	7.00
522636.51	572977.00	5.10
524657.77	566977.00	6.10
524657.77	568977.00	7.00
524657.77	570977.00	6.10
526657.00	564977.00	5.20
526657.00	568977.00	6.60

Table 1

The primary samples

We diffuse the information derived from this sample with the generator function g. If we consider 4 points $(x_k - \delta, y_k)$, $(x_k + \delta, y_k)$, $(x_k, y_k - \delta)$ and $(x_k, y_k + \delta)$, around on points (x_k, y_k) then the membership value of these points to fuzzy set F =proj$_{Oxy}(A)$ are

$$\mu_k = \exp\left(-\left(\frac{\delta}{\lambda}\right)^2\right) \tag{3}$$

where δ is the small distance between sample points and λ is the spread of the F.

The membership function of fuzzy set G =proj$_{OpH}(A)$ is

$$\mu_{pH,k} = \exp\left(-\left(\frac{\delta}{\lambda(pH)}\right)^2\right) \tag{4}$$

where $\lambda(pH)$ is the spread of the G.

Consequently, the derived sample is

$$\overline{A} = \left\{ (x_k, y_k, 1, pH_k, 1), (x_k \pm \delta, y_k, \mu_k, pH_k, \mu_{pH,k}), (x_k, y_k \pm \delta, \mu_k, pH_k, \mu_{pH,k}) / k = 1,...,n \right\} \tag{5}$$

where the first three components are input values and second two components are output values of BP network. We train the network with these samples. After the training process we get the estimation function of pH value in the region of the sample points. Using this estimation function we can calculate the pH value in any point of the studied region.

In this study we used a BP network with three layer: the first layer has 12 neurons and the activation function is tansig $(\text{tansig}(x) = 2/(1+\exp(-2*x))-1)$; the second layer has 12 neurons and the activation function is tansig; the last layer has 2 neurons and the activation function is purelin $(\text{purelin}(x) = x)$. For training we use the Levenberg-Marquardt method with regularization (Hagan and Menhaj, 1994) and we get the training parameters shown in Figure 3. As it can be seen the Sum of Square Errors (SSE) is 0.0463767.
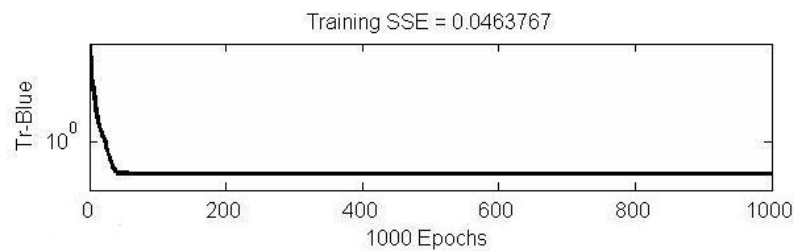

Figure 3

The training process with derived sample

In our application $\mu_{pH,k} = \mu_k = 0.96$. Figure 4 shows the membership values of the 24 samples after the training process.
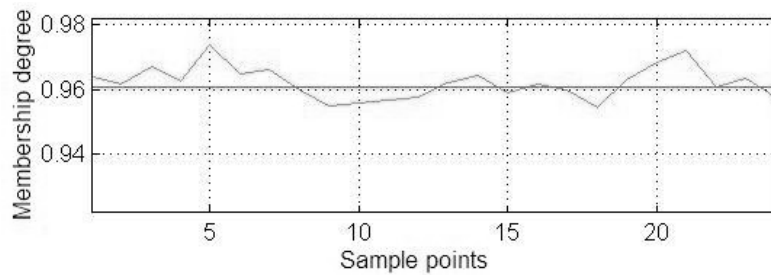

Figure 4

The membership values of samples

What can we see in Figure 4? The poligonal line shows the membership value of pH in the sample points. For example, in the 20th sample point the calculated pH value with trained network has a probability of 0.965. We can see that all membership value is greater than 0.95. This means, if we presume normal distribution, then the probabilty degree of estimation made from derived samples is 0.95. For verify this, we compared the contour lines of pH obtained with kriging (Isaaks and Srivastava, 1989) of 24 sample points to contour lines obtained from derived samples $\bar{A}$ using 14 primary samples A. We can see a good resemblance.
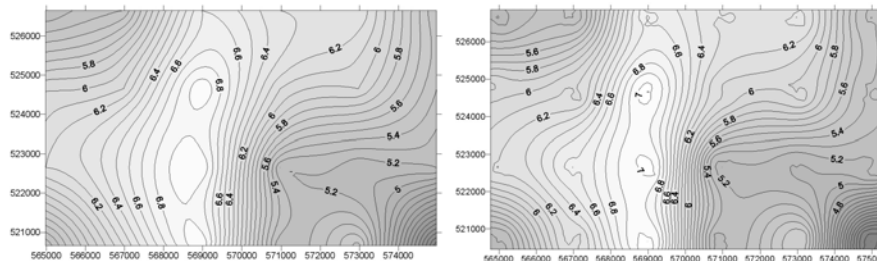
Figure 5

Contour lines obtained with kriging from 24 points and countour lines obtained from derived samples using 14 primary sample points

Conclusions

The difficulty of this method resides in selection of the diffusion function. In part 4 of this paper we presented a practical method for the calibration of the diffusion function. From the 24 known points we used 14 for pH values diffusion and the rest 10 for checking. Figure 5 (second picture) shows that it can be obtain a better aproximation from less sample points using diffusion function than with the use of primary sample points (Figure 2 – second picture).

Acknowledgement

References

[1]   G. J. Chaitin: Information, Randomness & Incompleteness, Word Scientific, Singapore, 1990

[2]   G. Cybenko: Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals, and Systems, 2(4) (1989), 303-314

[3]   M. T. Hagan and M. Menhaj: Training feedforward networks with the Marquardt algorithm, IEEE Transaction on Neural Networks, 5 (1994), 989-993

[4]   K. Hornik, M. Stinchcombe and H. White: Multilayer feedforward networks are universal approximators, Neural Networks, 2(5) (1989), 359-366

[5]   C. F. Huang and Y. Shi: Towards Efficient Fuzzy Information Processing – Using the Principle of Information Diffusion, pHysica-Verlag, Heidelberg, 2002

[6]   C. F. Huang and C. Moraga: A diffusion-neural-network for learning from small samples, Internat. J. of Approx. Reason. 35 (2004), 137-161

[7]   E. H. Isaaks, R. M. Srivastava: An Introduction to Applied Geostatistics. Oxford University Press, New York, Oxford, 1989

[8]   M. Kovács: A stable embedding of ill-posed linear systems into fuzzy systems, Fuzzy Sets and Systems 45 (1992) 305-312

[9]   Z. Makó: Quasi-triangular fuzzy numbers, theory and applications, Scientia, Cluj-Napoca, 2006

[10]   A. B. Mcbratney, B. Minasny, S. R. Cattle, R. W. Vervoort: From pedotransfer function to soil inference system. Geoderma 109 (2002), 41-73

[11]   Ya. A. Pachepsky, W. J. Rawlas: Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Sci. 63 (1999), 1748-1756

[12]   M. G. Schaap, F. L. Leij, M. Th. Van Genuchten: Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Sci. 62 (1998), 847-855

[13]   J. H. M Wösten, P. A. Finke, M. J. W. Jansen: Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66. (1995), 227-237

[14]   J. Wray and G. G. R. Green: Neural networks, approximation theory, and finite precision computation, Neural Networks, 8(1) (1995) 31-37