Comparison of regression pedotransfer functions and artificial neural networks for soil aggregate stability simulation

Adele Alijanpour shalmani[A], Mahmoud Shabanpour Shahrestani[B], Hossein Asadi[C] and Farid Bagheri [D]

[A] Msc. Student, University of Guilan, Rasht, Iran, Email adele.alijanpour@yahoo.com
[B]Faculty of Agriculture, University of Guilan, Rasht, Iran, Email m_shabanpur@yahoo.com
[C]Faculty of Agriculture, University of Guilan, Rasht, Iran, Email asadi@guilan.ac.ir
[D] Researcher, Tea Research center, Lahijan, Iran, Email faridbagheri@yahoo.com

Abstract
Simulation of soil aggregate stability is a suitable method for saving time and cost spent for direct measurement. This research comprises regression pedotransfer functions (RegPTFs) and artificial neural networks (ANNs) for estimation of soil aggregate stability. 140 soil samples from forest and rangland's soils of Guilan Province were collected and geometric mean diameter (GMD), %silt (Si), %clay (Cl), %sand (Sa), bulk density (BD), equivalent carbonate calcium ($CaCO_3$), particle density (PD), soil mechanical resistance (Load), pH, electrical conductivity (EC) and %organic matter (OM) values were determined. The data were split randomly into a calibration data subset (112 samples) and validation data subset (28 samples). Regression pedotransfer functions was performed by stepwise method and for establishing ANNs we used Marquardt-levenburg training algorithm and 3-layer perceptron structure with number of six neurons in one hidden layer. The best model of Regression functions for calibration GMD data was GMD=6.926-0.118pH-2.216PD-0.002Sa+0.103Load with R_{adj}^2=0.39. For determination of best ANNs model, we used five input patterns. Result showed that artificial neural networks with pH-PD-Sa-Load input pattern with R_{adj}^2= 0.87 for calibration GMD data, had most accurate prediction. With comparison of ANN with pH-PD-Sa-Load input pattern and regression pedotransfer functions, we found that ANNs with pH-PD-Sa-Load input pattern had higher R_{adj}^2 and Lower MSD (mean square of deviation) and hence ANNs could estimate soil aggregate stability better than regression pedotransfer functions.

Key Words
Simulation, soil aggregate stability, Pedotransfer function, artificial neural networks

Introduction
Soil aggregate stability determination is essential to erosion and conservation of soil, but direct measurement of Soil aggregate stability is time consuming and costly and so are called "Costly measured properties". However several researches have been done for indirect estimation of Soil aggregate stability from surrogate data such as texture, organic matter and bulk density. Regression pedotransfer functions and artificial neural networks are methods that can be used for simulation of Soil aggregate stability. Bouma (1989) expressed relationship between soil hydraulic properties and surrogate data such organic matter and bulk density and named it regression pedotransfer functions. Using regression pedotransfer function is not restricted to soil hydraulic properties estimation and used for simulation of soil chemical, biological and other physical properties. Artificial neural networks are intelligent modeling methods and can be used for costly measured soil properties estimation. They have the capability of learning complex relationship between multiple input and output variables (Nemes *et al*. 2002). Analysis of the ANN parameters suggested that more input variable and accurate data set were necessary to improve the prediction of costly measured soil properties (Tamari *et al*. 1996; Merdun *et al*. 2006).

Methods
In this research, 140 soil samples were collected from forest and rangland's soils of Guilan province. Soil samples were taken in each field at 0–20 cm depth for chemical and physical analyses. Then organic matter was determined by the Walkley and Black method, equivalent carbonate calcium was determined by titration method, pH was measured in suspension of soil to 0.01 M $CaCl_2$ ratio of 1:2.5 and electrical conductivity was measured in suspension of soil to water ratio of 1:5 (Page et al. 1982). Bulk density was determined by cylinder, particle density was determined by pycnometer, soil mechanical resistance was determined by penetrometer, fractions were used to measure particle size distributions (after complete dispersion with sodium hexametaphosphate) by the hydrometer method (klut. 1986) were determined as independent variables, and geometric mean diameter was determined by wet sieving apparatus (klut. 1986) was measured

as dependent variable. The data were split randomly into a calibration data subset (112 samples) and validation data subset (28 samples). Moreover, data subset used for determining the performance of two simulation method; artificial neural networks (ANNs) and regression pedotransfer functions (RegPTFs). Estimation of soil aggregate stability using RegPTFs were initially carried out using SPSS 14 for windows with stepwise method.

For establishing ANNs, We used Neural Works plus software with marquardt-levenburg training algorithm and 3-layer perceptron structure with number of six neurons in hidden layer. The number of neurons in the input and output layers corresponded to the number of Input and output variables. The number of hidden layers and its neurons is determined by try and error method and assumed equal to 1 and 6 respectively. Activation function was defined as a sigmoid tangent function. The performance of the PTFs estimating the soil aggregate stability, were assessed using two criteria: coefficient's statistics of corrected explanation (R_{adj}^2), mean square of deviation (MSD).

Results

Regression equation for estimation of calibration GMD data are showed in Table 1. Our postulate was the best model has the lowest MSD and the highest R_{adj}^2. Descriptive statistics for GMD using five ANN models and regressions pedotransfer functions are summarized in Table 2. Graphs for best model ANN and same pattern in $RegPTF_s$ calibration data subset for GMD estimation with input independent data pH-PD-Sa-Load are showed in Figure 1. The R_{adj}^2 values of both five ANN models and regression pedotransfer functions were significant based on the analysis of variance (ANOVA test) (P<0.01). Generally, both ANN and regression models could predict GMD accurately but ANN performed slightly better. Artificial neural networks are better than regression models for simulation soil aggregate stability (Mohammadi. 2002).

Table 1. Regression equation for estimation of GMD of calibration data

input independent variables	regression equation
Load-Sa-$CaCO_3$	**GMD**=0.526+0.109**Load**-0.005**Sa**-0.016 **$CaCO_3$**
Sa-PD-$CaCO_3$	**GMD**=9.419-0.004**Sa**-2.946**PD**-0.026 **$CaCO_3$**
pH-PD-Sa-Load	**GMD**=6.926-0.118**pH**-2.216**PD**-0.002**Sa**+0.103**Load**
pH-PD-Sa	**GMD**=10.041-0.144**pH**-2.945**PD**-0.003**Sa**
pH-PD-Si	**GMD**=9.935-0.142**pH**-3.008**PD**+0.004**Si**

Table 2. Descriptive statistics for GMD using 16 ANN models and regressions pedotransfer functions

input independent variables	R_{adj}^2 (cal) ANN	MSD (cal) ANN	R_{adj}^2 (cal) RegPTFs	MSD (cal) RegPTFs	R_{adj}^2 (test) ANN	MSD (test) ANN
Load-Sa-$CaCO_3$	0.85	0.066	0.34	0.288	0.77	0.129
Sa-PD-$CaCO_3$	0.77	0.101	0.22	0.371	0.28	0.407
Ph-PD-Sa-Load	0.87	0.058	0.39	0.265	0.57	0.245
pH-PD-Sa	0.61	0.172	0.17	0.371	0.09	0.516
pH-PD-Si	0.83	0.073	0.20	0.372	0.37	0.345

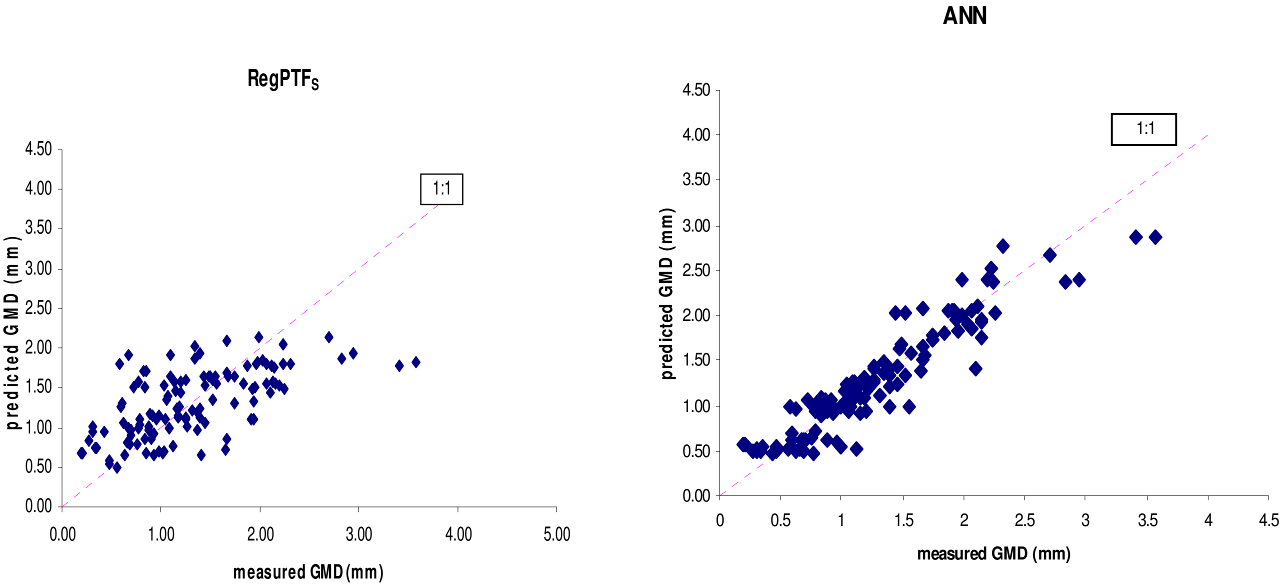

Figure 1. Graphs for best model ANN and similar pattern in $RegPTF_s$ calibration data subset for GMD estimation with input independent data (pH-PD-Sa-Load)

Conclusion

The best model of Regression functions for calibration GMD data was GMD=6.926-0.118pH-2.216PD-0.002Sa+0.103Load with R_{adj}^2=0.39 and MSD=0.265 for determination of best ANNs model, we used five input patterns. Result showed that artificial neural networks with pH-PD-Sa-Load input pattern with R_{adj}^2= 0.87 and MSD=0.058 for calibration GMD data, had most accurate prediction. With comparison of ANN with pH-PD-Sa-Load input pattern and regression pedotransfer functions, we found that ANNs with pH-PD-Sa-Load input pattern had higher R_{adj}^2 and Lower MSD (mean square of deviation) and hence ANNs could estimate soil aggregate stability better than $RegPTF_s$.

References

Bouma J (1989) Using soil survery data for quantitative land evaluation. *Soil Science Society of American Journal* **90**,177-213.

Klut A (1986) Method of Soil Analysis. Part 1.Physical and Mineralogical Properties. ASA, SSSA, Madison, Wisconsin, USA.

Merdun H, Cinar O, Meral R, Apan, M (2006) Comparioson of artificial neural network and regression pedotramsfer functions for prediction water retention and saturated hydraulic coundunctivity. *Soil & Tillage Res* **90**, 108-116.

Mohammadi J (2002) Testing an artificial neural network for predicting soil water retention characteristics from soil physical and chemical properties. 17 th WCSS. Thailand. Paper no:378. Paper no: 943.

Nemes A, Pachepsky Y, Rawls W, Wosten H, Zeilguer, A (2002) Using similarity and neural network approach to inter plate soil particle- size distribution. 17 th WCSS. Thailand. Paper No: 221. Paper No: 943.

Page AL, Miller RH, Keeney, DR (1982) Method of Soil Analysis. Part 2. Chemical and Microbiological Properties. ASA, SSSA, Madison, Wisconsin, USA.

Tamari S, Wosten JHM, Ruiz-Suarez, JC (1996) Testing an artificial neural network for predicting soil hydraulic conductivity. *Soil Science Society of American Journal* **60**, 1732-1741.