World Applied Sciences Journal 8 (9): 1065-1072, 2010
ISSN 1818-4952

Comparison of Regression Pedotransfer Functions and Artificial Neural Networks for Soil Aggregate Stability Simulation

[1]Adele Alijanpour shalmani, [2]Mahmoud Shabanpour Shahrestani, [2]Hossein Asadi and [3]Farid Bagheri

[1]Msc. Student, University of Guilan, Rasht
[2]Faculty of Agriculture, University of Guilan, Rasht, Iran
[3]Researcher, Tea Research center, Lahijan, Iran

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. 100 soil samples from forest and pasture's soils of Guilan Province in Iran 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 carbon (%) (OC) values were determined. The best model of regression functions for calibration GMD data was GMD=6.926-0.118pH-2.216PD-0.002Sa+0.103Load with R^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^2= 0.87 for calibration GMD data, had most accurate prediction. With comparison calibration GMD data 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^2 and lower RMSE and hence ANNs could estimate soil aggregate stability better than regression pedotransfer functions.

Key words: Simulation · Soil aggregate stability · Pedotransfer function · Artificial neural networks · Geometric mean diameter · Guilan

INTRODUCTION

Soil structure is critical for the germination and growth of plants and for the transport of water and contaminants through the unsaturated zone underlying agricultural fields. Soil structure may be defined as 'the spatial heterogeneity of the different components or properties of soil'. In other words, it is the variation of solids and voids as a function of scale that defines soil structure [1].

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 [2] 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 [3].

Artificial neural network is an attempt to build numerical techniques that are supposedly analogous to biological human neural system. Artificial neural network that were used in this research consist of an input, hidden and output layer, all containing simple autonomous processing elements (neuron, nodes, units) which are connected by adaptable communication paths called connectors [4]. Each connector is parameterized with a numeric value (weights) which indicated the strength of

Corresponding Author: Adele Alijanpour shalmani, MSc. Student, University of Guilan, Rasht, Iran.
Tel: +981317230458

the connection between the connected neurons and ability to pass signals. The number of neurons in input and output layers correspond to the number of input and output variables of the model. The number of hidden neurons can be varied freely but the optimal number depends on uncertainty and complexity of the modelling problem [3]. All input neurons j=1…J with the input variables $x_1...x_j$, are linked to all hidden layer neurons k=1…K by means of numeric adaptable connectors "weights" (Wjk). The input values is multiplied by weights and summed at the hidden neurons (Eq. 1). The hidden neurons consist of weighted input and bias (W_{j0}). A bias is simply a weight with constant input of 1 that serves as a constant added to the weights and these are calculated from a set of data through training process [5].

$$S_k = \sum_{j=0}^{J}(w_{jk}x_j) + w_{j0} \tag{1}$$

The result, S_k is used as a input for a So called activation function such as sigmoid functions yielding the hidden neuron output H_k (Eq 2).

$$H_k = \frac{1}{1+e^{-S_k}} \tag{2}$$

Then H_k are multiplied by the weights of W_{kl} (Eq 3) and in a same way as H_k, model outputs, Y_l are calculated (Eq 4).

$$Z_l = \sum_{k=0}^{k}(W_{kl} \times H_k) \tag{3}$$

$$Y_l = \frac{1}{1+e^{z_l}} \tag{4}$$

Artificial neural networks can be used for simulation soil particle size distribution [3], saturated hydraulic conductivity [4, 6, 7], Water retention curve properties [7, 8] and another soil properties such as soil loss and runoff [9, 10], soil dielectric constant [11] and nitrate-nitrogen in drainage water [12].

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 [6, 13].

MATERIALS AND METHODS

In this research, 100 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 carbon was determined by the Walkley and Black method, equivalent carbonate calcium determines from calcimetery 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 [14]. 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 [15] were determined as independent variables and geometric mean diameter was determined by wet sieving apparatus [15] was measured as dependent variable. The data were split randomly into a calibration data subset (80 samples) and validation data subset (20 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 number of neurons is determined by try and error method and assumed equal to 1 and 6 respectively. Activation function was defined as a sigmoidal tangent function. The performance of the PTFs estimating the soil aggregate stability, were assessed using three criteria: regression coefficient (R^2), root mean square of error ($RMSE$) and relative improvement (RI).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{N}(Y_i - \bar{Y}_i)^2} \tag{5}$$

Where;
Y_i- value of measured data,
$\hat{Y}$ - value of predicted data via model,
N- mean of measured data and
R^2- regression coefficient.

The root mean square of error, indicated mean accuracy of prediction which represents the expected magnitude of error (Eq 6). [5, 7].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \qquad (6)$$

Where;
Y_i- value of measured data,
$\hat{Y}$ - value of predicted data via model,
N- number of observations and
RMSE- root mean square of error.

And the another certain was relative improvement (*RI*) and related to performance improvement from one model (a) to another (b) (Eq 7) [5, 7].

$$RI = \frac{RMSE_a - RMSE_b}{RMSE_a} \qquad (7)$$

Where;

$RMSE_a$- root mean square of error for obligatory model a,
$RMSE_b$- root mean square of error for obligatory model b and *RI*- relative improvement.

R^2 and *RMSE* were calculated for calibration and validation data subset and assumed that the best model is model with highest R^2 and the lowest *RMSE*.

RESULTS AND DISCUSSION

Triangle of soil texture for soil samples in this research are showed in Figure 1. Regression equations for estimation of calibration GMD data are showed in Table 1. Our postulate was the best model has the lowest *RMSE* and the highest R^2. Descriptive statistics for GMD using five ANN models and regressions pedotransfer functions are summarized in Table 2. The R^2 values of both five ANN models and regression pedotransfer functions were significant based on the analysis of variance (ANOVA test) (P<0.01).

For soil aggregate stability estimation using $RegPTF_s$. In the best model of $RegPTF_{s,}$ pH, particle density, soil mechanical resistance and %sand entered in regression equation. Equation regression coefficient for this

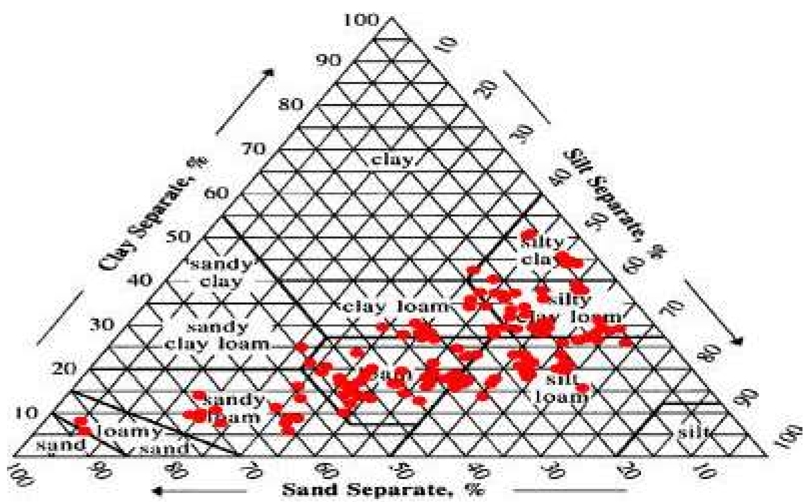

Fig. 1: Triangle of soil texture for soil samples

Table 1: Regression equations for estimation of GMD of calibration data

input independent variables	regression equation
Load-Sa-$CaCO_3$	GMD=0.526+0.109Load-0.005Sa-0.016 $CaCO_3$
Sa-PD-$CaCO_3$	GMD=9.419-0.004Sa-2.946PD-0.026 $CaCO_3$
pH-PD-Sa-Load	GMD=6.926-0.118pH-2.216PD-0.002Sa+0.103Load
pH-PD-Sa	GMD=10.041-0.144pH-2.945PD-0.003Sa
Load-Si-PD	GMD=5.801+ 0.105Load+0.005Si-2.166PD

Table 2: Descriptive statistics for GMD using five ANN models and regressions pedotransfer functions

input independent variables	$R^2_{(cal)}$ANN	*RMSE* $_{(cal)}$ANN	$R^2_{(cal)}$RegPTFs	*RMSE* $_{(cal)}$RegPTFs	*RI*	R^2(test)ANN	*RMSE* (test) ANN
Load-Sa-$CaCO_3$	0.85	0.257	0.34	0.537	77.08	0.77	0.359
Sa-PD-$CaCO_3$	0.77	0.318	0.22	0.609	72.78	0.28	0.638
pH-PD-Sa-Load	0.87	0.241	0.39	0.515	78.11	0.03	0.723
pH-PD-Sa	0.61	0.415	0.15	0.608	53.64	0.09	0.718
Load-Si-PD	0.79	0.305	0.37	0.522	65.93	0.20	0.670

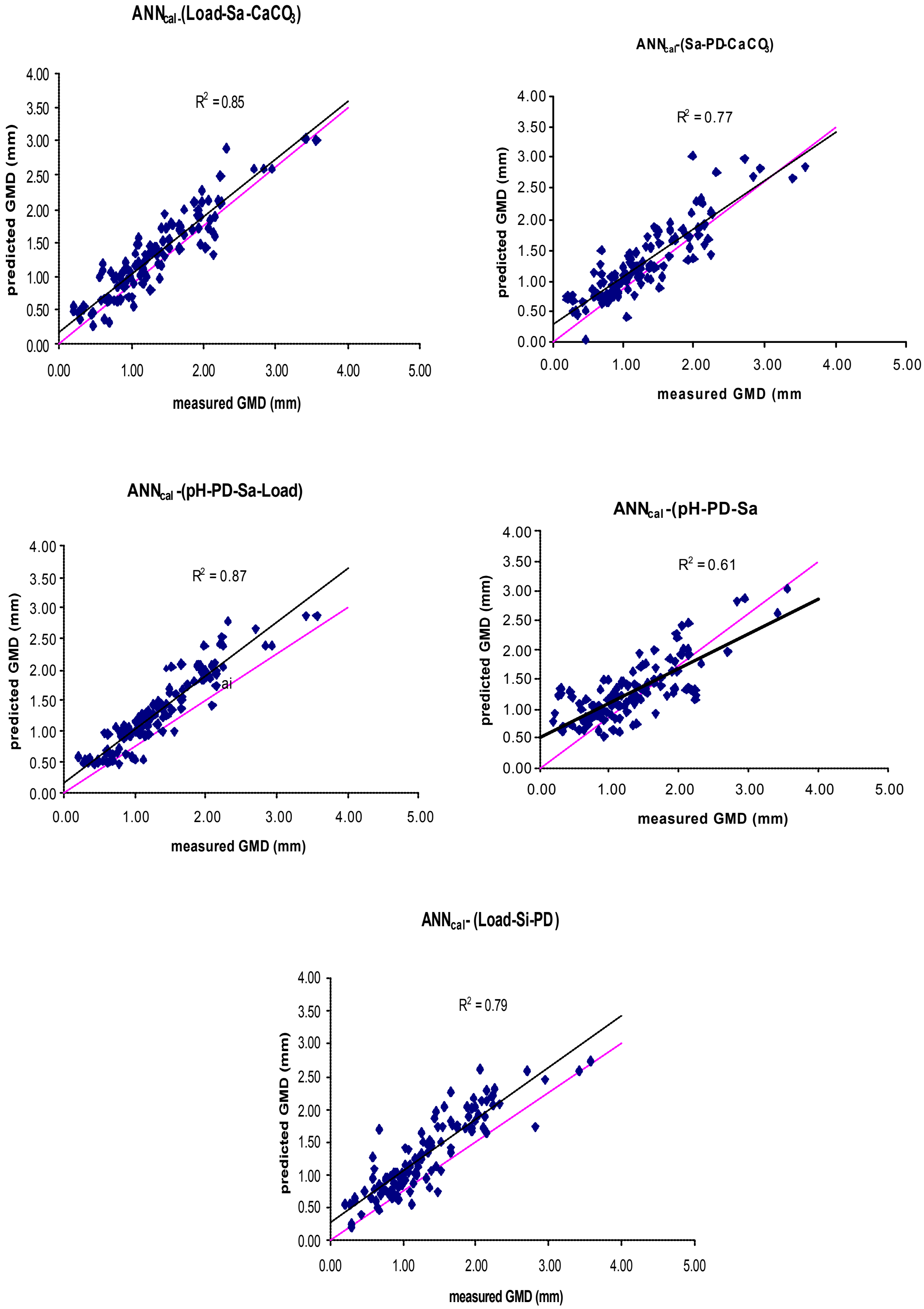

Fig. 2: Graphs for ANN models calibration data subset for GMD estimation

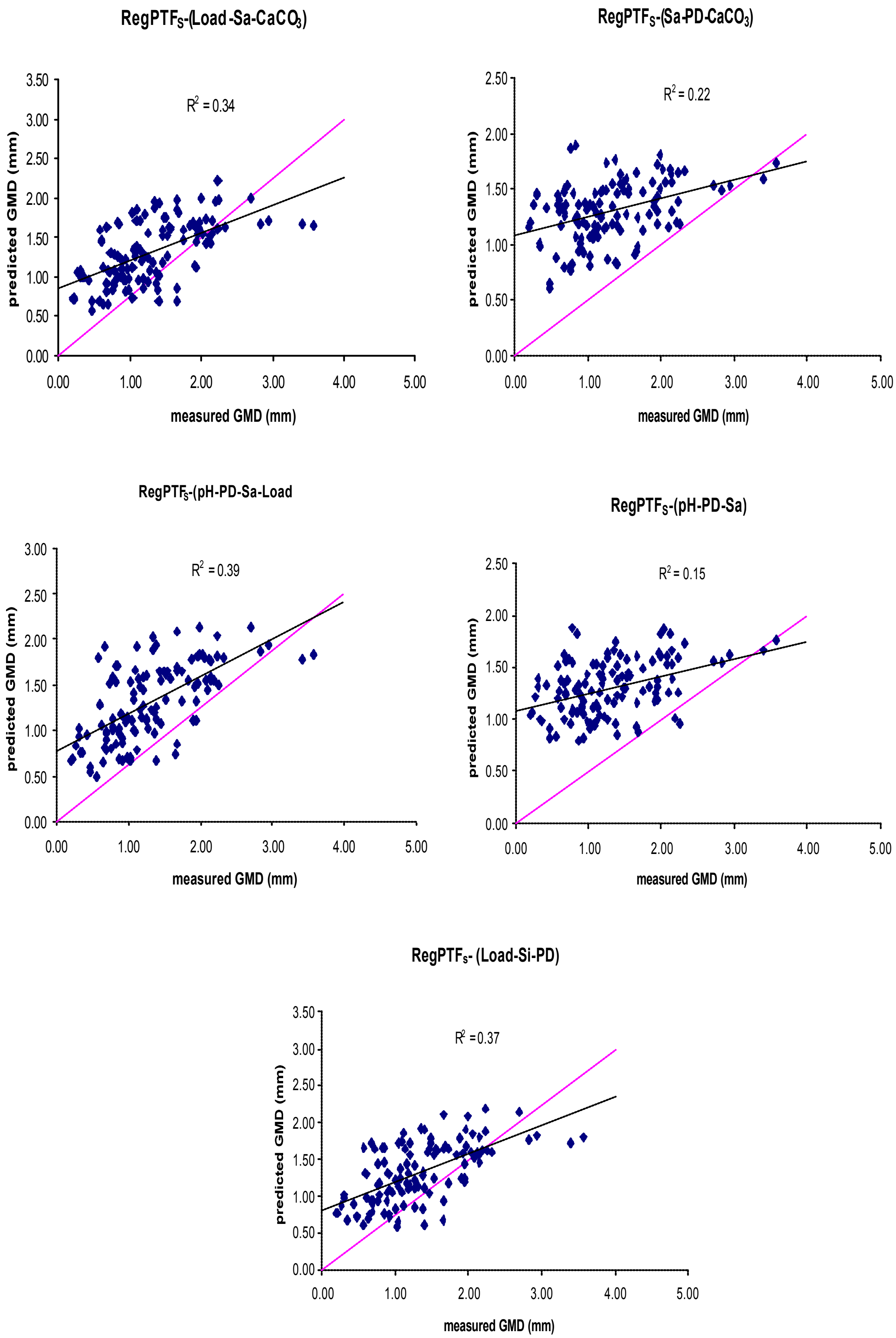

Fig. 3: Graphs for RegPTF$_s$ models calibration data subset for GMD estimati

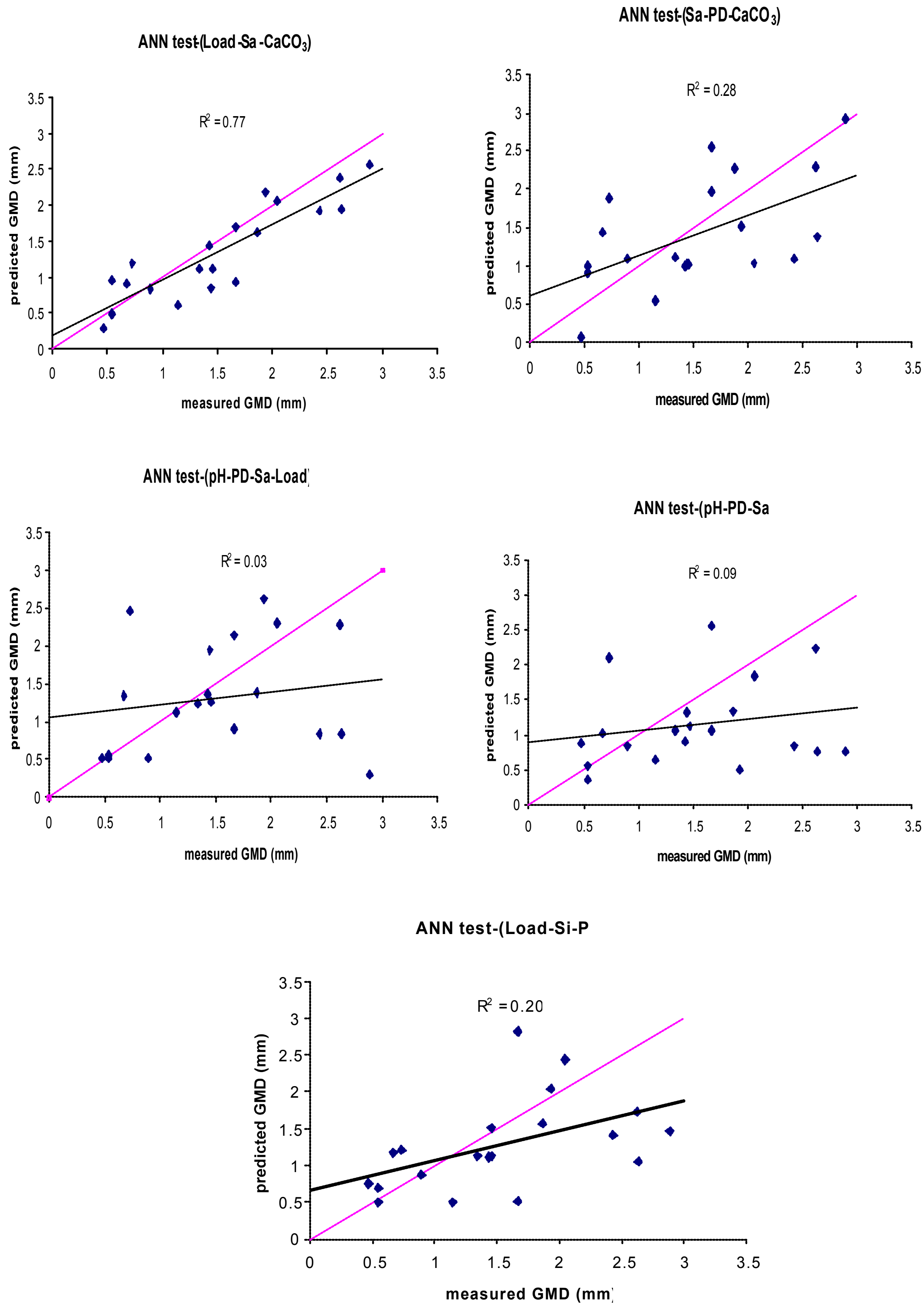

Fig. 4: Graphs for ANN models test data subset for GMD estimation

variables were -0.219, -0.345, 0.564 and -0.218 respectively and all of them were statistically significant ($P<0.01$) with R^2 = 0.39 and *RMSE* =0.515 (Eq 8).

GMD=6.926-0.118pH-2.216PD-0.002Sa+0.103Load (8)

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^2= 0.87 and *RMSE* =0.241 for calibration GMD data, had most accurate prediction. The best model of ANNs for test GMD data was Load-PD-$CaCO_3$ input pattern with R^2= 0.77 and *RMSE* =0.359. With comparison calibration GMD data 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^2 and lower *RMSE*. general comparison of R^2 and *RMSE* of artificial neural network and regression pedotransfer functions for estimation of calibration GMD data showed that artificial neural networks had a higher R^2 and lower *RMSE* in compared with regressions pedotransfer functions, hence artificial neural networks predicted percent more than the changes in the data that this results and Ghielmiand Eccel [16], Heuvelmans *et al.* [17] and Tamary *et al.* [6] is similar.

Also, with comparison calibration GMD data of ANN and $RegPTF_s$, we found that all of the ANN models had higher *RI* than $RegPTF_s$ and hence we saw that the ability of artificial neural networks in estimation GMD changes, is more than the ability of linear regression pedotransfer functions.

Graphs for ANN models calibration data subset for GMD estimation showed in Figure 2, Graphs for $RegPTF_s$ models calibration data subset for GMD estimation showed in figure 3 and Graphs for ANN models test data subset for GMD estimation showed in Figure 4.

In the graphs whatever points near the 1:1 line indicated to less deviation of predicted data than measured data and more accurate estimation of the model. Thus the graphs of artificial neural network model (Figure 2) with similar pattern in regressions pedotransfer functions (Figure 3) for estimation of calibration GMD data is shown, the values estimated by the artificial neural network had less diversion than regressions pedotransfer functions.

This is important to mention that the model test in the diagnosis relations between data that not observation previously is working and is necessarily the best model of calibration data, the best model of test data not be. For example, the minimum value of *RMSE* and maximum value of R^2 in artificial neural networks for estimation of calibration GMD data was related to model pH-PD-Sa-Load but the minimum value of *RMSE* and maximum value of R^2 in artificial neural networks for estimation of test GMD data was related to model Load-Sa-$CaCO_3$.

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 [18].

This species can be concluded that the ability of artificial neural network for estimation of calibration GMD data was further than regression pedotransfer functions. More accurate estimate of artificial neural network than regression pedotransfer functions because mode operation can be considered three layers perceptron network, network performance is to be that first pattern presented to network and its output is calculated, then the network comparison predicted output values with measured output variable or variables and coefficients weight of network were kind of change that the lowest variation is between predicted data and measured data [4].

REFERENCES

1. Pirmoradian1, N., A.R. Sepaskhah1 and M.A. Hajabbasi, 2005. Application of Fractal Theory to quantify Soil Aggregate Stability as influenced by Tillage Treatments. Biosystems Engineering, 90: 227-234.
2. Bouma, J., 1989. Using soil survery data for quantitative land evaluation. Soil Sci. Society of American J., 90: 177-213.
3. Nemes, A., Y. Pachepsky, W. Rawls, H. Wosten and A. Zeilguer, 2002. Using similarity and neural network approach to inter plate soil particle- size distribution. 17th WCSS. Thailand. Paper No: 221. Paper No: 943.
4. Minasny, B., J.W. Hopman, T. Harter, S.O. Eching, A. Toli and M.A. Denton, 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multi step outflow data. Soil Science Society of American Journal, 68: 417- 429.
5. Minasny, B. and A.B. Mcbartney, 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci. Society of American J., 66: 352-361.
6. Tamari, S., J.H.M. Wosten and J.C. Ruiz-Suarez, 1996. Testing an artificial neural network for predicting soil hydraulic conductivity. Soil Sci. Society of American J., 60: 1732-1741.

7. Schaap, M.G. and F.J. Leij, 1998. Using neural networks to predict soil water retention and hydraulic conductivity. Soil and Tillage Res., 47: 37-42.
8. Matula, S., M. Mojrova and K. Spongrova, 2007. Estimation of the Soil Water Retention Curve (SWRC) Using Pedotransfer Functions (PTFs). Soil and Water Res., 2(4): 113-122.
9. Licznar, P. and M.A. Nearing, 2003. Artificial neural networks for soil erosion and runoff prediction at the plot scale. Catena, 51: 89-114.
10. Rosa, D., F. Mayol, J.A. Moreno, T. Bonson and S. Lozano, 1999. An expert system /neural network model (ImpelERO) for evaluating agricultural soil erosion in Andalucian region, southern Spain. Agri. Ecosys. Enviromen., 73: 211-226.
11. Persson, M., B. Sivakumar, R. Berndtsson, O.H. Jacobsen and P. Schjonning, 2002. Predicticting the dielectric constant-water content relationship using artificial neural network. Soil Sci. Society of American J., 66: 1427-1429.
12. Sharma, V., S.C. Negi, R.P. Rudra and S. Yang, 2003. Neural networks for predicting nitrate-nitrogen in drainage water. Agri. Water Management, 63: 169-183.
13. Merdun, H., O. Cinar, R. Meral and M. Apan, 2006. Comparioson of artificial neural network and regression pedotramsfer functions for prediction water retention and saturated hydraulic counductivity. Soil and Tillage Res, 90: 108-116.
14. Page, A.L., R.H. Miller and D.R. Keeney, 1982. Method of Soil Analysis. Part 2. Chemical and Microbiological Properties. ASA, SSSA, Madison, Wisconsin, USA.
15. Klut, A., 1986. Method of Soil Analysis. Part 1.Physical and Mineralogical Properties. ASA, SSSA, Madison, Wisconsin, USA.
16. Ghielmi, L. and E. Eccel, 2006. Descriptive models and artificial neural networks for spring frost prediction in an agricultural mountain area.. Computers and Electronics in Agriculture, 54: 101-114.
17. Heuvelmans, G., B. Muys and J. Feyen, 2006. Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets. J. Hydrol., (319): 245-265.
18. 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.