International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:4, No:11, 2010

Sensitivity Analysis for Determining Priority of Factors Controlling SOC Content in Semiarid Condition of West of Iran

Y. Parvizi, M. Gorji, M.H. Mahdian, M. Omid

***Abstract*—**Soil organic carbon (SOC) plays a key role in soil fertility, hydrology, contaminants control and acts as a sink or source of terrestrial carbon content that can affect the concentration of atmospheric CO2. SOC supports the sustainability and quality of ecosystems, especially in semi-arid region. This study was conducted to determine relative importance of 13 different exploratory climatic, soil and geometric factors on the SOC contents in one of the semiarid watershed zones in Iran. Two methods canonical discriminate analysis (CDA) and feed-forward back propagation neural networks were used to predict SOC. Stepwise regression and sensitivity analysis were performed to identify relative importance of exploratory variables. Results from sensitivity analysis showed that 7-2-1 neural networks and 5 inputs in CDA models output have highest predictive ability that explains %70 and %65 of SOC variability. Since neural network models outperformed CDA model, it should be preferred for estimating SOC.

***Keywords*—**Soil organic carbon, modeling, neural networks, CDA.

I. Introduction

SOIL organic carbon (SOC) has an indispensable role in the ecosystems and acts as a buffer to global climatic change. It is therefore critical to maintain its quality for the sustainability of ecosystems [30]. Soil organic carbon is a vital component, as it plays a key role in soil fertility, hydrology, contaminants control and acts as a sink or source of terrestrial carbon content that can affect the concentration of atmospheric CO2. The terrestrial carbon reservoir is estimated to be between 3,400 to 3,500 Giga tons that SOC consists about 3,000 Giga tons of them [25]. Hence, soil can be considered as important sink and source for carbon sequestration and modifier of climate changes [6]. Appropriate management of SOC can substantially decrease the atmospheric carbon that has increased exponentially at a rate of 1.5% per year [7]-[15]. Also, soil is regarded as an important sink and source for carbon sequestration and modifier of climate changes. This can be done by management and land use activities [27].

Y. Parvizi is with the Agriculture and Natural Resource Research Centre of Kermanshah, Iran (corresponding author to provide phone: 0098-831-8370070; fax: 0098-8318351022; e-mail: yparvizi@ut.ac.ir).

M. Gorji Assistant Professor Tehran University, Iran, (e-mail: mgorji@ut.ac.ir)

M.H. Mahdian ,Associate Professor, Agriculture Research and Education Organization, Iran. (e-mail: mahdian_1338@yahoo.com).

M. Omid Associate Professor Tehran University, Iran. (e-mail: omid@ut.ac.ir)

One approach in dealing with linear statistical methods in soil process modeling and SOC class estimation is not new.

These methods include multiple linear regressions (MLR), logistic regression, and canonical discriminate analysis (CDA).

However, there are characteristics of the models such as over simplification, ignorance of complex nonlinear interactions etc., which limit their use in accurately assessing the distribution of the C across the landscapes.

Another approach in dealing with nonlinear systems is to use artificial Intelligence (AI) modeling paradigm such as artificial neural networks (ANNs). ANNs has been successfully used in classification, prediction, and pattern recognition problems [8]-[14]-[31]. The potential benefits of ANNs include greater prediction reliability, cost-efficient estimation and solving complex problems involving nonlinearity and uncertainty. ANNs are inspired by biological neural networks [2]-[29]. ANNs learn from training examples, adjusting weights to reduce the error between the measured and the predictions.

ANNs have been successfully applied in various soil studies [26]. These applications include predictions of soil structure [17], pedotransfer functions [1]-[16]-[21, pedometric use in soil survey [13], environmental correlation of three-dimensional soil spatial variability [19], and prediction of SOC from soil parameters [10]-[11].

This study was conducted to model SOC variability at watershed scale across agricultural rainfed land use types in a semi-arid condition of Iran. We employed ANNs and CDA to investigate the effect of climatic, topographic and soil properties and also management variables on SOC. Various ANNs topologies were designed and tested. Since the estimation model was data based, selection of appropriate input and output variables is important. Thus, sensitivity analysis on exploratory variables was carried out to select the best input combinations for modeling and estimating SOC.

II. Materials and Methods

Site Description

Merek watershed from Karkheh river basin with area about 24200 ha, was selected for this study, because in this watershed we can find appropriate diversity in soil, topography and semi-arid climatic conditions. The elevation of this site ranged from 1450 to 19174. It has a semi-arid and cold climate with an annual precipitation of about 500 mm. The land use is mainly: agricultural land that covers about 14500 ha. In this site, soil texture is clay or salty clay and soil structure is blocky. Soils, in mountains and highlands, were covered by about 25-60%

fine and coarse gravel. PH is 7.3-7.9, soil salinity is generally low (0.4-0.8 ds/m), lime content in topsoil is 4-60%.

Sampling and Dataset

A sampling design based on stratified random sampling method was performed taking into account land use, soil, slope and aspect map. In each sampling points, soil samples were taken from topsoil (0–0.3 m depth) in three different landuses. Figure 1 illustrates the sampling scheme and sample positions. In total 245 soil samples were collected, air dried and then sieved by 2 and 0.5 mm sieve. Soil organic carbon (SOC) of the samples was determined using colorimetric method. Some soil physiochemical properties including; calcareous content (%TNV), sand, silt and clay contents and saturation percent were determined based on standard laboratory methods.

Apart from soil properties, climatic variables included mean annual temperature (MAT), mean annual rainfall (MAR), potential evapotranspiration (ETP) and climate types in each sampling point were recorded from Amberger climate map that was generated based on 50 years climatic data from climatic stations in the region [3]. Topographic variables (including elevation, slope and aspect of the terrain of the sampling locations) were also measured in sampling time using clinometer and compass, respectively. Geometric factors (such as, curvature, and terrain parameter) were derived from DEM that prepared based on digitized contour line map with 20 meter vertical lag apart. The transformed aspect (TA), which aligns the index along a SW-NE axis, for the sites was calculated using the following equation [5]:

$$TA = \cos(45 - aspect) \quad (1)$$

TAP parameter was calculated by multiplying TA by sinus value of slope angle. This parameter was used to incorporate the effects of slope on direct-beam radiation.

Development of ANNs

A typical ANN consists of interconnected processing elements included: an input layer, one or more hidden layers, and an output layer (which provides the answer to the presented pattern). Between input and output layers there could be several other hidden layers . The input layer contains the input variables for the network while output layer contains the desired output system, and the hidden layer often consists of a series of neurons associated with transfer functions. The propagation of data through the network starts with the presentation of an input stimulus at the input layer. The data then progress through, and are operated on by the network until an output stimulus is produced at the output layer.

Data Selection and Preprocessing

Selection of input data was based on theoretical contribution of physical variable, expert experiences and accessibility. Initially, 15 inputs were entered to the input layer of networks. Inputs were selected from climatic factors, soil physical properties, geometric factors and landuse types. The effect of landuse type was involved with a classified variable in three class (1 = agriculture lands, 2 = forest and 3 = rang).

The training algorithm of this network was GDM. For network training, cross validation was implemented as the stopping criteria. The data set was split into a training set and a testing set. The trained model was validated with the testing set sequentially. The training was terminated when the prediction error of the testing dipped into a minimum and started to increase. Before simulation, all data sets were standardized by the software using a linear algorithm. In the present study, a three-layer feed forward back propagation ANN was used. First, the number of nodes in the hidden layer was optimized.

Sensitivity Analysis

Since the SOC estimation model was data based, selection of appropriate input and output variables is important. A sensitivity test was performed on the chosen ANN's so that a better understanding of the relative importance of each input on the output could be examined. Thus, sensitivity analysis was carried out to investigate the dynamic behavior of input variables such as MAT, AR, EVT, soil physical properties, etc. This was done by imposing steps changes to various inputs and observing their effects on the network output. These responses were used as guides to select appropriate input and output variables that are suitable for model development.

Canonical Discriminate Analysis (CDA)

The CDA analysis was performed on the data that was already used to develop the neural networks. These data first classified in three classes (low, medium and high SOC contents) by univariate clustering method. Model was linearly developed by all 15 exploratory variables. In the second model, stepwise CDA was used to develop a model for predicting SOC classes. A validation dataset was used to validate the CDA models, whereas test dataset was used to test the performances of the CDA equations.

Performance Criteria

The performance of the models was evaluated by a set of test data using mean square error (MSE), coefficient of determination (R^2) on testing set, between the predicted values and the target (or experimental) values as follows [17]:

$$MSE = \frac{1}{n}\sum_{i=n}^{n}\left(SOC_i^{*} - SOC_i\right)^2 \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(SOC^{*}{}_i - SOC_i)^2}{\sum_{i=1}^{n}(SOC_i - \overline{SOC})^2} \quad (13)$$

Where SOC^{*}_i is the network (predicted) output from observation i, SOC_i is the measured SOC from observation i, $\overline{SOC}$ is the average value of measured output, and n is the total number of data observation. For performance evaluation

of CDA criteria, we used confusion table or confusion matrix and diffusion chart of canonical discriminate functions (CDF). Additionally, we used in this study two different measures including the Mean Bias Error (MBE) and the correlation coefficient (ρ) between different exploratory variables and SOC values and also between predicted and measured values by models. The MBE is a measure of bias and reveals the overestimation or underestimation.

$$MBE = \frac{1}{n}\sum_{i=n}^{n}\left(SOC_i^{*} - SOC_i\right) \quad (14)$$

III. Results and Discussion

Statistical characteristic of selected variables across watershed are summarized in appendix table. In general, selected variables are strongly heterogeneous as indicated by their coefficient of variation (CV). For example, slope varies from 1 to 46% with CV of more than 90%. While, mean annual temperature and rainfall of the study region show a small variation, indicating the climatic variation within the samples. The SOC content varied from 0.37 for Dry lands with abandoned erosion to 3.31% in mountain range soils**.** The correlation coefficients (ρ) SOC and soil TNV, SP and climatic variable (MAT, AR, ETP and climate type) were significant ($\alpha = 0.05$). But ρ-value between SOC and aspect components, curvature and sand percent was not significant. As expected, the correlations between the SOC and TNV, silt, MAT and ETP were all negative.

Prediction of SOC by CDA

First canonical discriminate function can explain 63.8% of variance between three classes that explored by predicted variable. Second CDF found 36.2% of variability among the means. Scatter plot in fig. 1 explain distribution of CDF amounts that calculated by CDF method. Distance among center classes explains ability of method to determine SOC level by exploratory variables.

Confusion matrix of outputs evaluates CDA ability to determine SOC level by 15 exploratory variables (see Table 1). Results indicated that CDA method can correctly predict about 705 of SOC levels in samples. Ability of CDFs to predict SOC levels in high level of SOC was grater and in medium level is lower than others.

Relative importance of exploratory variables (see Table 2). In this table parameters was explain the relative weights of the variable to predict SOC levels. This parameters reflect the relative importances of each variable to predict three level of SOC. Relative importance of exploratory variables to determine low SOC level were MAT, Climate type, TA, curvature and TA.

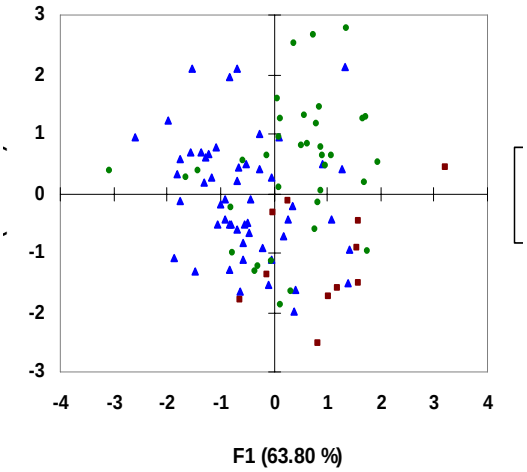

Fig. 1 Distribution plot of CDFs

TABLE I
Confusion Matrix for the Estimation Sample

from \ to	1	2	3	Total	% correct
1	40	9	9	58	68.97 %
2	10	72	6	43	62.79 %
3	2	1	9	12	75.00 %
Total	52	73	24	113	67.26 %

To locate samples in high and medium SOC level, we must consider above variables. Therefore these variables respectively have highest importance to determine SOC levels in this climate condition from agricultural land use type.

TABLE II
Parameters of Classification Functions in CDFs:

	SOC classes		
	1	2	3
Intercept	-3710.65	-3744.99	-3708.53
elevation	2.06	2.06	2.05
slope	-3.39	-3.36	-3.32
TA	-33.58	-33.17	-32.27
TAP	12.05	11.23	8.64
curvature	-14.81	-21.36	-22.23
T.N.V	0.78	0.76	0.68
S.P	5.25	5.39	5.53
Gravel	-0.16	-0.11	-0.16
clay	-1.39	-1.40	-1.44
silt	-1.20	-1.20	-1.35
sand	0.00	0.00	0.00
M.A.T	198.00	198.85	197.41
A.R	2.41	2.44	2.42
ETP	-0.04	-0.04	-0.04
ClimateType	146.54	146.25	148.32

ANN Structure

Finding the optimum number of hidden neurons in the hidden layer is an important step in developing MLP networks. In neural network design, too many hidden units cause over-fitting, while too few hidden units cause underfitting. A

summary of our findings and best networks architecture is shown in Table 3. Among the different configurations examined, the 15–9–1 configuration, exhibited highest accuracy. The performance of this network is shown in Fig. 2. The MSE values for the 15-*k*-1 ANN's, with different *k* values and epoch's, presented that when k=9 in training set and k=2 in calibration and validation data sets the model was not overt trained and optimum epoch in validation set equal to 358 (see Table 3). To find the optimum number of hidden units, the MSE of the network with 15 inputs were plotted against the number of hidden units (Fig. 2).

TABLE III
ERROR INDICES AND BEST STRUCTURE FOR DIFFERENT DATA SET OF ANN

	Train	CV	Test	Best Networks	Train	CV	Topology
MSE	0.121	0.083	0.120	Hidden PEs	9	2	15-2-1
MAE	0.199	0.305	0.275	Epoch #	594	376	
MBE	0.03	0.02	-0.1	Final MSE	0.0123	0.0413	
ρ	0.651	0.376	0.631				

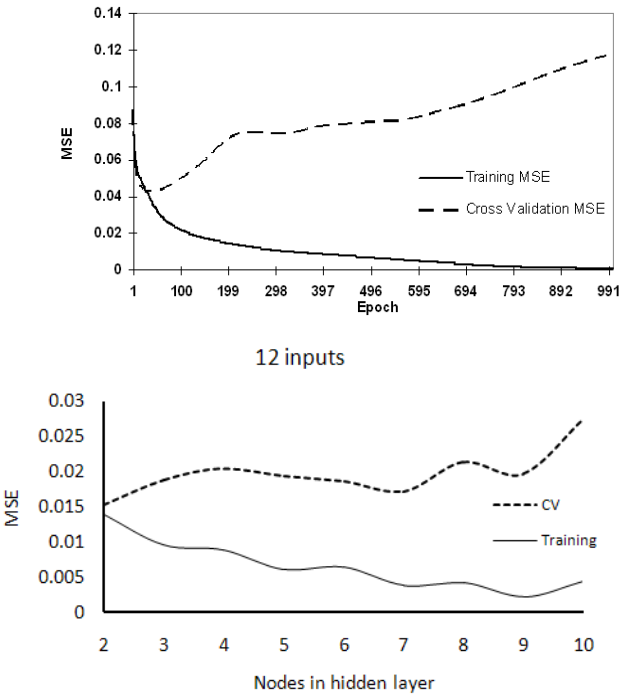

Fig. 2 MSE of different epochs (below) and nodes in hidden layers (above) in CV and Training data sets

SensitivityAanalysis

In order to test the hypothesis that not all of the inputs used were required to train the model networks effectively, it was necessary to measure the influence of each input variable on the output. This was done by measuring the mean rate of change of each output when a single input was changed by some relatively small amount (0.001). The mean rate of change was determined by testing the model network 594 times using randomly selected input values. Sensitivities, defined as the MSE rate of change outputs divided by rate of change of a given input. Fig 3 indicated sensitivity analysis results.

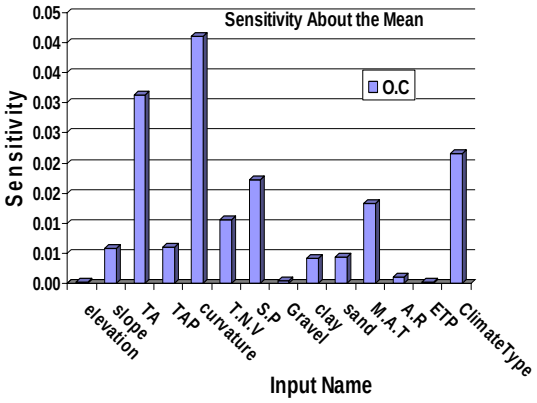

Fig. 3 Sensitivity analysis of the exploratory variables

This analysis showed relative importance of exploratory variable on SOC variability prediction. Result indicated that curvature, TA, climate type, SP, MAT, TNV, slope and integrated slope-aspect influence on net solar radiation that reflected in TAP variable, have highest impacts on this variability, respectively. The application of best methods to estimate SOC from climatic, soil and geometric factors could account for only about 76% of the variations in this study (see figure 4). Several factors could be implicated for this rather modest accountability. Among them are the kind and status of soil management practices (Stewart and Hossner, 2001).

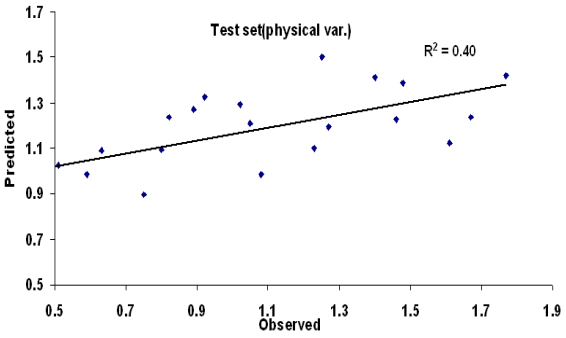

Fig. 4 Plot of predicted vs observed SOC in ANN methods

IV. CONCLUSIONS

In this study, artificial neural networks and CDA approaches were employed to develop models for predicting SOC change in a semi-arid condition in west of Iran. Available climate,

topographic and soil properties variables data were used. Sensitivity analysis and parameters CDF comparison were also done to determine relative importance of these variable components in ANN and CDA, respectively.

The newly developed neural network with 9 neurons in the hidden layer predicted SOC better than the regression models and improved the accuracy of the prediction. The ANN models are in general more suitable for capturing the non-linearity of the relationship between variables. In this study, however, the relationship between SOC and exploratory variables appeared to be dominantly linear. The intelligent models (ANNs) could also derive climatic factors influence and integrated effects of slope-aspect factors on net solar radiation, that reflected in TAP index, on SOC variability The models tested in this study, using physically based variables included: curvature, TAP, TNV, Climate type, SP, MAT and AR could account for only up to 69% of the variation in SOC in the semiarid conditions of Iran.

Analysis of sensitivity accuracy in ANN showed that adding more variables such as elevation, clay, silt and ETP can only improved 7% variability prediction and did not significantly improve the modeling results. Results of ANN sensitivity analysis and coefficient of variables in CDA model showed that climate and geometric factors has highest impact of SOC variability in agricultural land use type. It seemed we must brought to our attention that to improve predictability power of our methods considering management factors specially tillage, straw and grazing management could be attempted. We hope to include these in our future works.

ACKNOWLEDGMENTS

The financial support provided by the University of Tehran and agricultural research organization of Iran is gratefully acknowledged.

REFERENCES

[1] M. Amini, K.C.Abbaspour, H.vKhademi, N.vFathianpour, M. Afyuni,, R. Schulin, *Neural network models to predict cation exchange capacity in arid regions of Iran*. Europ. J. of Soil Sci. vol. 56, pp. 551–559, 2005.

[2] J.A. Anderson.. *Introduction to neural networks.* Prentice-Hall of India, New Delhi. 2001

[3] APERI, *Mahidasht-Sanjabi plain study :phase 1: climate study. TAM consulting engineers*, Ministry of Agriculture, Iran.) vol. 2, 2004.

[4] T.W. Beers, Dress, P.E., Wensel, *Aspect transformation in site productivity research.* J. of Forertry, vol. 64, pp. 691–692. 1966.

[5] C.E.P. Cerri, M. Easter, K. Paustian, K. Killian, K. Coleman, M. Bernoux, P. Falloon, D.S. Powlson, N.H. Batjes, E. Milne, C.C. Cerri, *Predicted soil organic carbon stocks and changes in the Brazilian Amazon between 2000 and 2030*. Agric., Ecosys. and Environ. Vol. 122, pp. 58–72, 2007.

[6] IPCC, *Land-use, land-use change, and forestry*. In: Watson, R.T., Noble, I.R., Bolin, B., Ravindranath, N.H., Verardo, D.J., Dokken, D.J. (Eds.), A Special Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge. 2000.

[7] M. Kim, T. Kim, *A neural classifier with fraud density map for effective credit card fraud detection.* IDEAL, pp. 378–383. 2002.

[8] Lal, R.,. *Soil carbon dynamics in cropland and rangeland. Environ.* Pollut. Vol. 116, pp. 353–362, 2002.

[9] E.R. Levine, D.S. Kimes, *Predicting soil carbon in Mollisols using neural networks*. pp. 473–484. In R. Lal et al. (ed.) Soil process and the carbon cycle, CRC Press, Boca Raton, FL. 1997.

[10] E.R. Levine, D.S. Kimes, V.G. Sigilitto, *Modeling soil structure using artificial neural network*. Ecol. Modell. Vol. 92, pp. 101–108, 1996.

[11] D. Liu, Z. Wang, B. Zhang, K. Song, X. Li, J. Li, F. Li, H. Duan, *Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China*. Agric., Ecosys. and Environ. Vol. 113, pp. 73–81, 2006.

[12] A.B. McBratney, B. Minasny, S.R. Cattle, R.W. Vervoot, *From pedotransfer function to soil inference system.* Geoderma vol. 109, pp. 41–73, 2002.

[13] J. McCullagh, *A modular neural network architecture for rainfall estimation, Artificial Intelligence and Applications.* Innsbruck, Austria, pp. 767–772, 2005.

[14] K. McVay, C. Rice, *Soil organic carbon and the global carbon cycle*. Technical Report MF-2548, Kansas State University, Kansas. 2002.

[15] B. Minasny, A.B. McBratney, *The neuron-m method for fitting neural-network parametric pedotransfer functions*. Soil Sci. Soc. Am. J. vol. 66, pp. 352–361, 2002.

[16] A. Nemes, M.G. Shaap, J.H. Wo, M. Sten, *Functional evaluation of pedotransfer functions derived from different scales of data collection.* Soil Sci. Soc. Am. J. vol. 67, pp. 1093–1193, 2003.

[17] M. Omid, A. Baharlooei, H. Ahmadi, *Modeling Drying Kinetics of Pistachio Nuts with Multilayer Feed-Forward Neural Network*. Drying Tech. vol. 27, no.10, pp. 1–9, 2009.

[18] S.J. Park, P.L.G. Vlek, Environmental correlation of three dimensional soil spatial variability: A comparison of three adaptive. Geoderma , vol. 109, pp. 117–140, 2002.

[19] F. Sarmadian, R. Taghizadeh Mehrjardi, A. Akbarzadeh, *Modeling of some soil properties using artificial neural network and multivariate regression in Gorgan province, north of Iran*. Austral. J. of Basic and Applied Sci. vol. 3, no. 1, 323-329, 2009.

[20] S. Somaratne, G. Seneviratne, U. Coomaraswamy, *Prediction of Soil Organic Carbon across Different Land-use Patterns:A Neural Network Approach*. Soil Sci. Soc. Am. J. vol. 69, pp. 1580–1589, 2005.

[21] G.P. Sparling, D. Wheeler, E.T. Wesely, L.A. Schipper, *What is soil organic matter worth?*. J. Environ. Qual. Vol. 35, pp. 548–557, 2006.

[22] M.J. Spencer, T. Whitfort, J. McCullagh, Dynamic ensemble approach for estimating organic carbon using computational intelligence. *Proceedings of the 2nd IASTED international conference on Advances in computer science and technology*. Puerto Vallarta, Mexico, 2006.

[23] Z. Tan, R. Lal, *Carbon Sequestration Potential Estimates with Changes in Land Use and Tillage Practice in Ohio*, USA. Agric., Ecosys. and Environ., vol. 126, pp. 113-121, 2005.

[24] Z. Tan, R. Lal, N. Smeck, F. Calhoun, *Relationships between surface soil organic carbon pool and site variables*, Geoderma, vol. 121, pp. 187–195, 2004.

[25] P.J. Werbos, *Backpropagation, basic and new developments.* pp. 134-139. In Arbib, M.A., (ed.) The handbook of brain behavior and neural networks. The MIT Press, Cambridge, MA. 1995.

[26] B. Wolf, G.H. Snyder, *Sustainable Soils: the place of organic matter in sustaining soil and their productivity*. Food Products Press. New York, 2003.

[27] G. Zhang, *Neural Networks in Business Forecasting*, IRM Press, Hershey, PA., 2004.