ECOPERSIA

DOI:10.18869/modares.Ecopersia.5.1.1699 **2017, 5 (1): 1699-1709**

Intelligent Approaches to Analysing the Importance of Land Use Management in Soil Carbon Stock in a Semiarid Ecosystem, West of Iran

Yahya Parvizi[1]* and Mosayeb Heshmati[1] and Mohammad Gheituri[1]

[1] *Assistant Professor, Department of Soil Conservation and Watershed Management, Agriculture and Natural Resource Research Center of Kermanshah, AREEO, Kermanshah, Iran*

Received: 5 October 2016 / Accepted: 12 December 2016 / Published Online: 20 March 2017

ABSTRACT The effects of different climatic, soil, geometric, and management factors on soil organic carbon (SOC) degradation and sequestration potential was evaluated in the semi-arid zone of Mereg watershed, west of Iran. Two nonparametric methods, viz. Classification and Regression Tree (CART) and feed forward back propagation Artificial Neural Network (ANN) were compared with parametric Multivariate Linear Regression (MLR) in estimation of SOC content. Soil sampling was conducted using randomized systematic method in work unit map by overlying soil, aspect and slope maps. Results indicated that linear models had higher prediction errors. The CART with all variables (physical and management) and the ANN with 31-2-1 topology carried the highest predictive capability, explaining 81% and 76% of SOC variability, respectively. ANN models overestimated SOC content and showed a higher capability to detect the effects of management factors on SOC variations. In all the methods, management factors dominantly controlled SOC stock sequestration or degradation in different land use.

Keywords: *CART, MLR, Neural networks, Semi arid environment, Soil organic carbon*

1 INTRODUCTION

Soil is an important resource that fundamentally supports the sustainability of life in ecosystems, while also acting as a buffer to global climatic change (Sparling *et al*., 2006; Attaeian, 2016). Soil organic carbon (SOC) stock acts as a sink or source of terrestrial C, affecting the concentration of atmospheric CO_2 and playing most important role to mitigate climate changes. These roles can be managed through proper land use activities (Tan and Lal, 2005; Sharma *et al*., 2014). Soils in semiarid conditions can be considered as more efficient source to sequester atmospheric CO_2 and, therefore, in mitigating climate change (Lal, 2008). Theoretically, SOC variability can be controlled by climatic and geometrical factors as well as soil type. In dry and sub-humid conditions, Han *et al*. (2009) indicated that SOC variability was mostly controlled by geometric variables especially slope gradient and aspect. Mismanagement and land use change was found to decrease 49% of rudimentary SOC in arid and semiarid conditions (Evrendilek *et al*., 2009). Heshmati *et al*. (2015) revealed that SOC spatial variability was strongly influenced by land use change and management agents as compared

Corresponding author: Department of Soil Conservation and Watershed Management, Agriculture and Natural Resource Research Center of Kermanshah, Kermanshah, Iran. Tel: +98 833 838 3460, E-mail: yparvizi1360@gmail.com

with physical factors. Therefore, it is important to delineate and quantify the effects of different physical (soil, geometric and climatic factors) and management factors along with their interactions on SOC variability, especially in semiarid conditions.

Application of such statistical methods as MLR to delineate the effects of physical and management factors on SOC variability were prevailed. To avoid some limitations of these methods, including over simplification, ignorance of complex nonlinear interactions, such nonlinear systems as CART and ANN that use nonparametric methods can be employed (Zhang, 2004; Mc Cullagh, 2005). The potential benefits of these methods are greater reliability of prediction and solving complex problems involving nonlinearity and uncertainty (Spencer *et al.*, 2006). They have been successfully applied to predict pedotransfer functions (Amini *et al*., 2005; Sarmadian *et al*., 2009), pedometric use (Mc Bratney *et al*., 2002), as well as environmental correlation of soil spatial variability (Park and Vlek, 2002).

This study was conducted to quantify relative importance of factors controlling SOC variability leading to atmospheric carbon sequestration in soil at watershed scale across rainfed, forest and range lands in semiarid environments of Iran. Applied methods included MLR method as linear and parametric approach along with ANN, and CART methods as nonlinear nonparametric approaches. Stepwise elimination (in MLR algorithm) and sensitivity analysis (in ANN method) on 31 exploratory variables were carried out. These techniques were applied to determine the relative importance of physical and management variables for controlling SOC stock variability in semiarid environments.

2 MATERIALS AND METHODS

2.1 Site description

The study site was situated 693800, 694600E and 3769700-3770600N in the Mereg watershed of Kermanshah province, west of Iran (Figure 1), with an elevation range of 1450 -1850 m, cold semi-arid and mean annual precipitation of about 500 mm. Soil temperature and moisture regimes are Mesic and Xeric, respectively (APERI, 2004). Soils texture ranges from clay to silt covered with about 25-60% of fine to coarse gravel in highlands. The pH varied between 7.3-7.9, EC 0.4 - 0.8 (ds m^{-1}), and 4-60% of lime content in topsoil. The site covers about 14500 ha of rainfed crop land, dominantly under wheat and pea rotation (APERI, 2004).

2.2 Soil sampling and data set

A randomized systematic sampling design was used on work unit map based on soil classification, slope, and aspect maps. Finally, 245 strata were separated after preparing work unit map and in total 199 soil samples were taken from topsoil (0–30 cm) after elimination of similar work unit in field work. Situation of sampling points across the land cover map plotted in Figure 1. Soil samples, collected in the designated land uses (forest, range and agriculture), were air dried before measuring their organic carbon contents in laboratory. Some other soil physicochemical properties were also determined, including percentages of total neutralizable value (TNV) by titration with normal NaOH, sand, silt, and clay by hydrometric methods, and saturation percent (SP) (Nelson *et al.*, 1996).

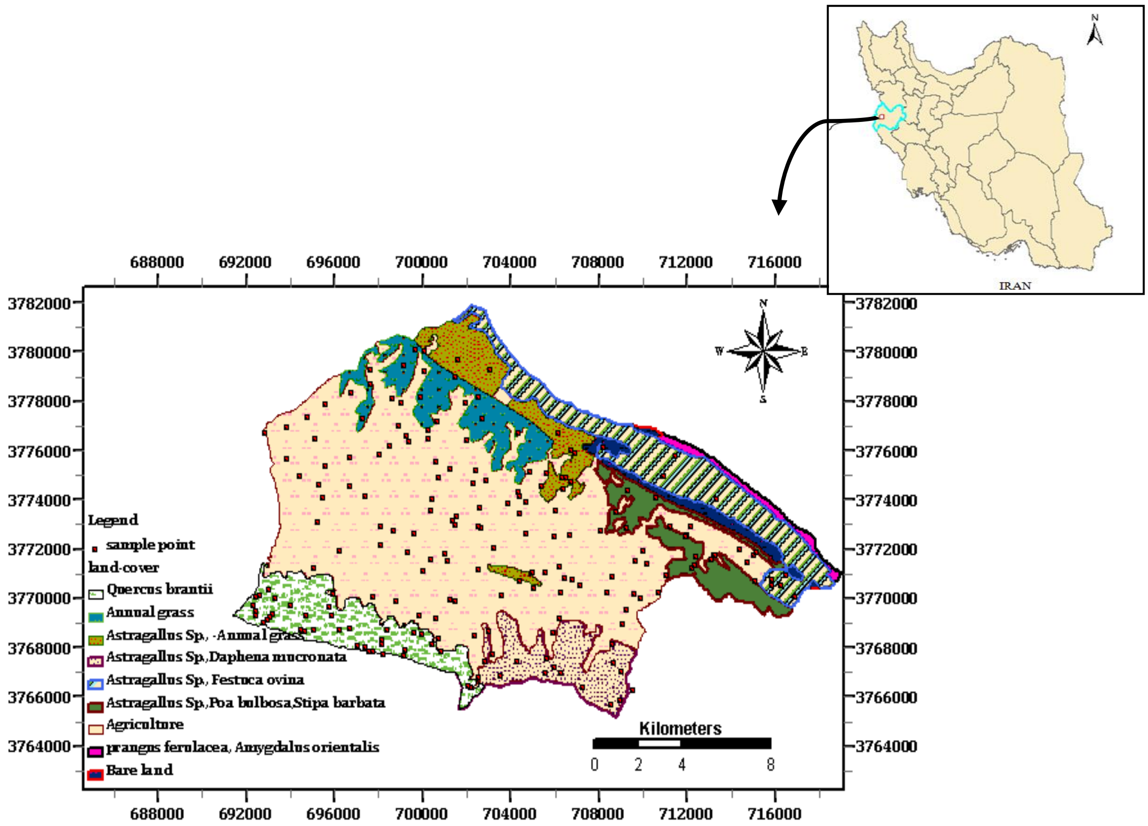

Figure 1 Site position and sampling points across land cover map of Mereg watershed

Also, some climatic variables, including mean annual temperature (MAT) and rainfall (MAR), potential evapotranspiration (ETP), and climate types (Ctype) were determined, using the Amberger method (APERI, 2004). Topographic variables, including elevation (Elev), slope (P), and aspect of the sampling site terrain were also determined. Such geometric factors as curvature (Curv), and terrain parameter were derived from 1:50000 Digital Elevation Model (DEM), prepared as based upon digitized contour line map of 20 meter vertical lag apart. The transformed aspect (TA), which aligns the index along a SW-NE axis, was calculated for sampling points according to Beers *et al.* (1966), using the following equation:

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

TAP parameter was calculated by multiplying TA by the sine value of slope angle. This parameter was employed to incorporate the effects of slope on direct-beam radiation (Beers *et al.* 1966). To investigate the management effects, using collected data from field study, 13 raw quantified and three combined sets of scenario indices, as management variables, were defined. The 13 primary variables included field size in ha, as ownership index (Oh), mannuring (Mn), legume and cereal frequency in rotations (L.F and C.F), prevalence of winter fallow (Wf), crop residue grazing (Gc), straw harvesting (Sh), burning of straw (Sb), domestic density (Dd), machinery energy consumption (E) (Mj ha^{-1} year^{-1}), tillage index (T_{index}), plough direction (P_{dir}), and finally accelerated soil degradation class (Er). Crop residual related variables (Gc, Sh, and Sb) were selected to define crop residue management scenario (Ssen) variables by clustering them through K-means clustering. Also tillage management scenario index (Tsen) was defined by classifying tillage related variables (E, T_{index}, and P_{dir}) with K-means clustering. Third management scenario variable was made by combining rotation related variables (Wf, L.F, and C.F) with K-means clustering. This variable is defined as rotation management system (Rsen) variable. Before processing the algorithm, the data set was split into training set (60%), testing set (20%) and cross validation set (20%).

2.3 Multiple Linear Regressions (MLR)

Two MLR equations were constructed using XLSTAT software. The first model was linearly developed by all 31 exploratory variables. In the second model, Stepwise Regression (SR) was applied to develop a regression model for predicting SOC (Sarmadian *et al.*, 2009). All the data sets were randomly divided three series including training, validation and test series. Validation data set was used to validate the MLR models, whereas test data set was applied to test the performances of the MLR equations.

2.4 Development of ANNs

A typical ANN consists of interconnected processing elements, including an input layer, one or more hidden layers, and an output layer which provides the answer to the presented pattern (Demuth *et al*., 2009). The input layer contains the input variables for the network while the output layer containing the desired output system and the hidden layer often consisting of a series of neurons associated with transfer functions. The total error at the output layer is distributed back to the ANN and the connection weights being adjusted. This process of feed-forward mechanism and back propagation (BP) of errors and weight adjustment is repeated iteratively until convergence in terms of an acceptable level of error is achieved (Krenker *et al*., 2011). In this study, gradient descent with momentum (GDM) algorithm was used for speeding up BP using Neuralsolution software.

2.5 Classification and Regression Trees (CART)

The CART represents a unification methodology of all tree-based classification and prediction methods. It transforms the regression tree models in a conspicuous nonparametric alternative to the classical methods of regression (Breiman *et al.* 1984). The CART algorithm creates a set of questions that consist of all possible questions concerning the measured variables. Then splitting criterion was done by maximum likelihood, creating a tree with one node containing all the training data using XLSTAT software. To avoid overtraining, pruning the tree was made through V-fold cross-validation (Spencer *et al*., 2006).

The best split is chosen to maximize a splitting criterion. When the impurity measure for a node can be defined, the splitting criterion corresponds to a decrease in impurity. Least-square deviation (R(t)) was used as the measure of impurity of a node that is computed as:

$$R(t) = \frac{1}{N_w(t)} \sum_{i \in t} w_i f_i (y_i - \bar{y}(t))^2 \qquad (2)$$

Where Nw (t) is the weighted number of cases in node t; wi is the value of the weighting variable for case i; fi is the value of the frequency variable; yi is the value of the response variable; and y(t) is the weighted mean for node t. Stopping rules control was: if node becomes pure; that is, all cases in a node have identical values of the dependent variable, the node will not be split.

2.6 Performance criteria and software

To evaluate the accuracy of the prediction models, the performance of the models were evaluated by set of test data using mean square error (MSE), root mean square error (RMSE), coefficient of determination (R^2) on testing set, between the predicted values and the target (experimental) values. In addition, the mean bias error (MBE) and the correlation coefficient (ρ) were taken into account. MBE is a measure of bias revealing either the overestimation or underestimation .

To establish various ANN's, a software package, NeuroSolutions (Version 5.02) was used. The expression used to calculate MSE is given by NeuroSolutions for Excel. CART algorithm and MLR were carried out through SPSS 16 along with XLSTAT pro-7.5 package.

3 RESULTS AND DISCUSSION

The SOC content varied from 0.34% for land with abundant erosion (main kind of visible erosion, including sheet, rill and gully erosion), to 3.72% in the soils received manure in agricultural land use. Prediction of soil carbon variation in corresponding predicting soil, geometric, climatic and management factors using different simulating data mining methods explained in following paragraphs.

The analysis of variance of the MLR model of SOC indicated that both MLR and SR models were highly significant (P<0.001, Table 1). The MLR model, which explains all the exploratory variables, and stepwise elimination involvement, respectively, explained 64 and 49 percent of SOC variations in the semiarid conditions (Table 1). Spencer *et al.* (2006) findings revealed that MLR model with physical combination in input variables could predict utmost 29-54% of SOC variability. Stepwise elimination model (Eq.3) indicated that TNV among physical, and Burn, T_{index}, Er., and Mn. among the management factors, linearly and significantly determined 49% of SOC variability in the rainfed crop lands in the semiarid conditions.

$$SOC = 1.923 - 0.01TNV + 0.28Burn - 0.71T_{index} - 0.09Er. \quad (3)$$

3.1 SOC simulated through MLR and SR

Table 1 Model summary, error index and analysis of variance of MLR model with all variables and after stepwise elimination (SR)

Model	RMSE	R^2	Sum of Squares	df	Mean Square	F	Sig.
MLR	0.385	0.643	30.2884	12	2.52404	20.563	0.000
SR	0.418	0.632	29.7482	5	5.9496	49.362	0.000

MLR: Predictors: (Constant) all 31 physical and management variables, SR: Predictors: (Constant) TNV, Mn., Burn., T_{index}, Er. Dependent variable: SOC

Table 2 Evaluation indices of nonparametric models with different input variables combination

Input variables	Nonparametric method	RMSE	MBE
All variables	CART	0.056	0.000
	ANN	0.026	0.025
Management variables	CART	0.106	0.000
	ANN	0.306	0.010
Physical variables	CART	0.155	0.000
	ANN	0.345	0.003

RMSE: root mean square of error, MBE: Mean bias error

3.2 SOC estimation through CART algorithm

Three different combinations of exploratory variables were applied for estimation of SOC contents through the CART algorithm. The CART model while using all the input variables with an exploration of about 80% of SOC variability had the highest efficiency in SOC estimation (Figure 2). This model indicated that management variables, as compared with physical agents (soil, geometric and climatic factors), more profoundly influenced SOC variability, and was able to identify 63 percent of SOC variability through a source of management factors. Application of the CART algorithm did not reveal any bias error in any combination of the predicting variables. But error estimations stood between 0.06 to 0.15% that can be neglected considering SOC range (0.37 – 3.72%) (Table 2). Therefore, CART algorithm could be considered as a good method to estimate SOC contents and as well to determine the management effects on it. CART algorithm was predominantly applied in estimating categorical soil variables while there werescarce scientific references in using it in continuous soil variables including SOC.

3.3 ANN's structure optimization

Theoretically, too many hidden units in training process cause overfitting, on the other hand reduced hidden units cause underfitting. Among the different tested configurations, network with 31–2–1 topology (31, 2 and 1 neurons in input, hidden and output layer, respectively), with tan-hyperbolic transfer function in hidden layer exhibited highest accuracy and least error on cross validation data set (MSE=0.0768) (Table 3). Somaratne *et al*. (2009) findings showed that tangent -sigmoid transfer function in hidden layer was a more suitable one.

The above optimum feature has 31 variables as input vector, 2 neurons in its hidden layer, and 1 neuron as output vector as shown in Figures 2 and 3. After evaluating the optimized configuration with the test data set, the MSE of 0.107, 0.113 and 0.120 were obtained when inputs included all, management, and physical variables, respectively. The corresponding-values were 0.88, 0.83, and 0.63. This input combination was able to significantly increase the predictive ability of ANN in comparison to Spencer *et al.* (2006) findings (with=0.59). The MSE values for the ANN's, with different nodes in hidden layers and epoch's, showed that when 2 nodes were in hidden layer in validation data sets, the model was not overtrained. Optimum epochs in validation set were 46, 358, and 376 in models were run applying all, management, and physical variables as input vector, respectively (Table 3).

Table 3 ANNs performance indices with best architecture in train, test and cross validation data set

Inputs		Train	CV	Test	Network attributes	Train	CV
	MSE	0.001	0.118	0.107	Hidden 1 PEs	9	2
All variables					Epoch #	1000	46
	ρ	0.860	0.862	0.883	Final MSE	0.001	0.043
Management variables	MSE	0.094	0.077	0.113	Hidden 1 PEs	10	2
					Epoch #	5000	358
	ρ	0.808	0.814	0.831	Final MSE	0.0018	0.0737
	MSE	0.121	0.083	0.120	Hidden 1 PEs	9	2
Physical variables					Epoch #	594	376
	ρ	0.651	0.376	0.631	Final MSE	0.0123	0.0413

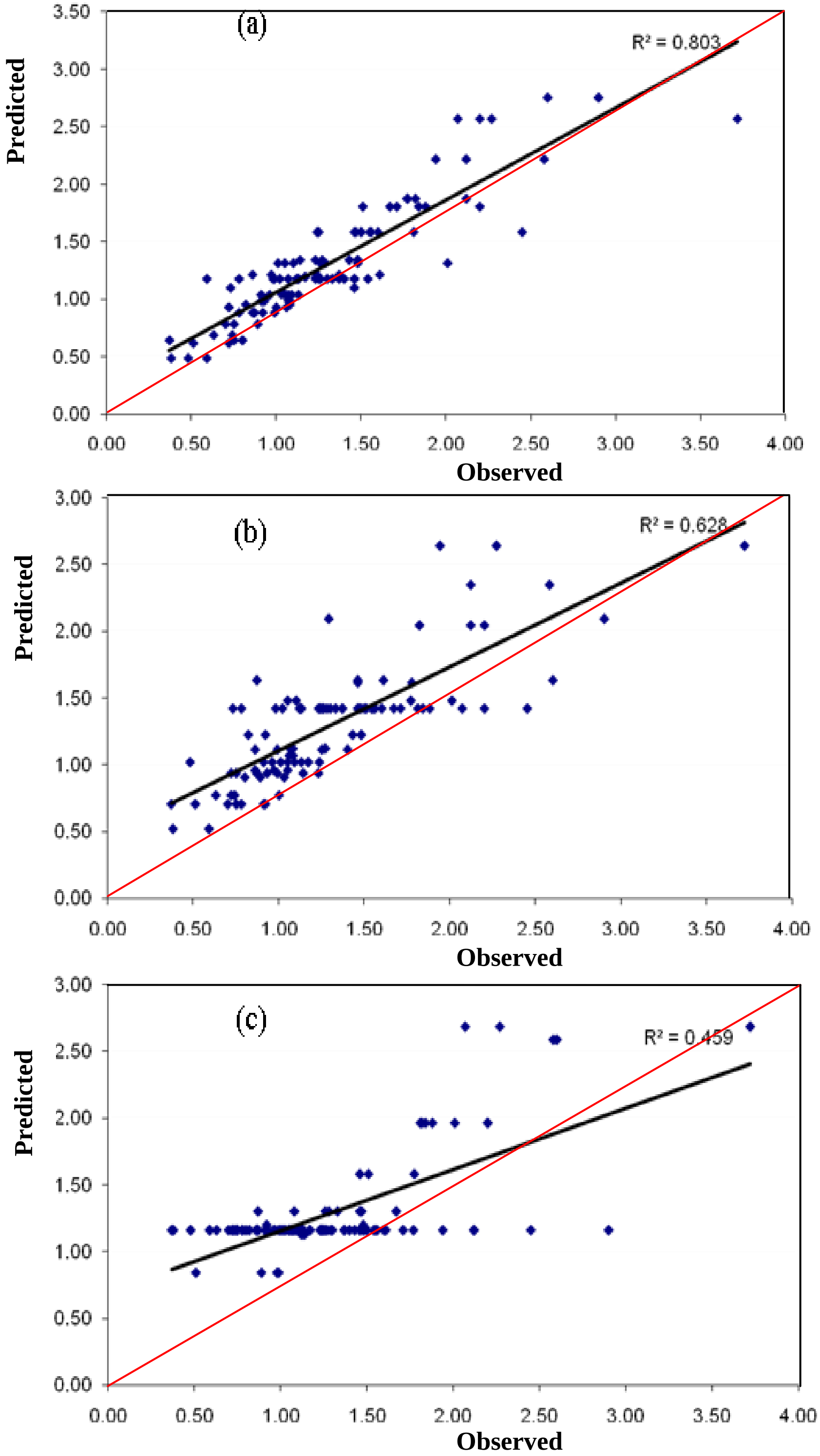

Figure 2 Estimated vs. measured SOC by CART models with all (a), management (b) and physical (c) variables as predictors

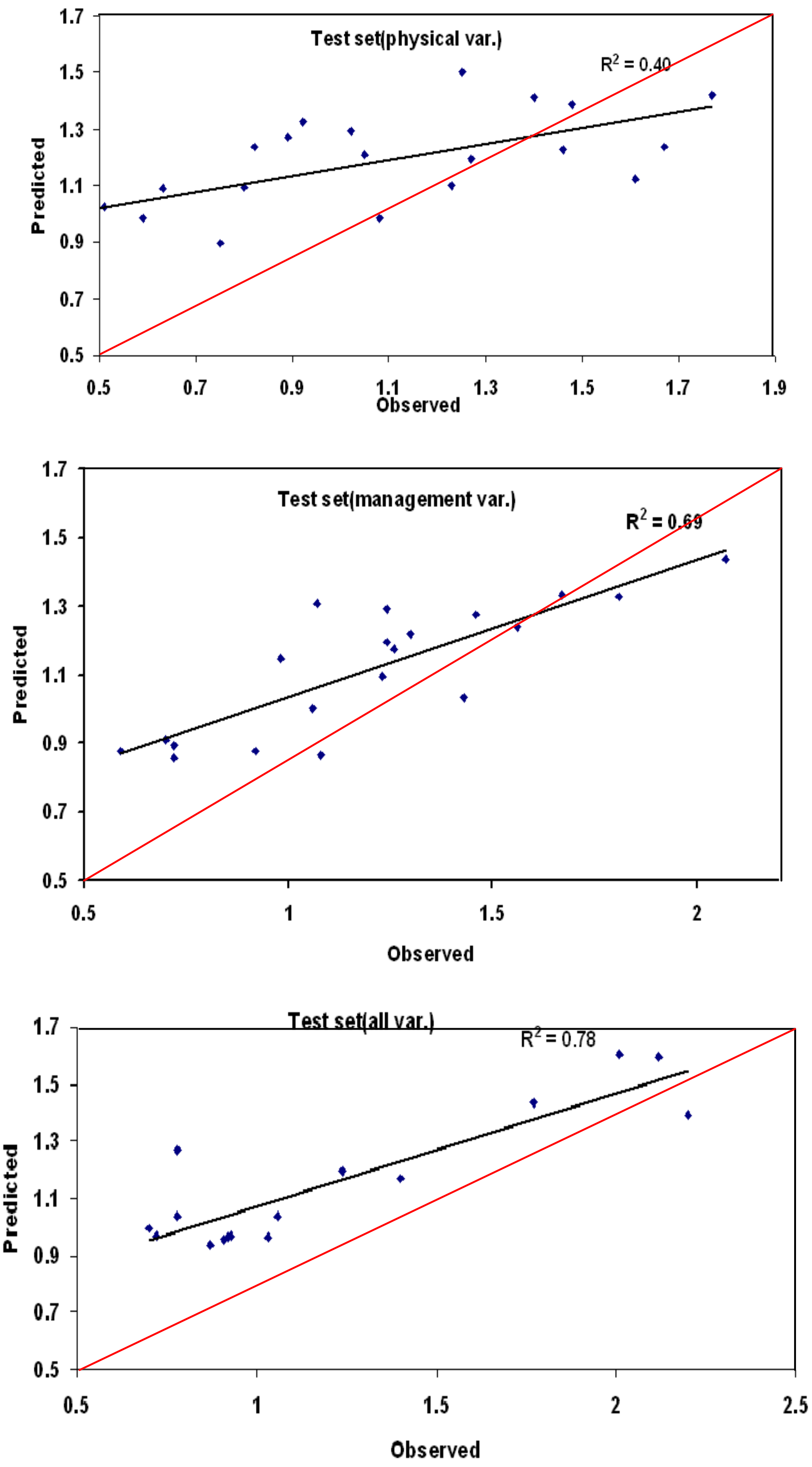

Figure 3 The scatter plot of the measured vs. estimated SOC using the ANNs with different inputs combination

The scatter plot of the measured against predicted SOC, in the test data set, is given in Figure 3 for the ANN models, which was identified as being the best model for predicting soil organic carbon.

4 CONCLUSIONS

Higher estimation error of parametric linear methods is the disadvantage of these methods comparing with nonparametric nonlinear methods. Findings of this research indicated that newly developed ANN could detect management agent effects on SOC stock variability more efficiently than the CART models. But, CART models could explore nonlinearity and interaction between variables more accurately than ANNs. Management factors especially tillage and crop residue scenario parameters, and also rotation parameters, predominantly determined SOC stock variability in rainfed land use in the semiarid conditions of the experiment. The nonparametric tested models (CART and ANN), using physically based variables including TAP, TNV, gravel, SP, MAT and AR, could account for only up to 40-45% of the variation of SOC stock in the study area.

For prioritizing the importance of variables to determine SOC stock variation, sensitivity analysis results revealed adding more other physical variables could slightly improve the prediction. But, no significant improving was evident in the modeling results of soil carbon stock and sequestration potential. It is recommended that in the future research, management factors especially tillage, rotation, straw, and grazing management could be more attentively taken into account, as it could improve the predictability power of our research methods.

5 ACKNOWLEDGMENT

The financial support provided by the ANRRC of Kermanshah Province is gratefully acknowledged.

6 REFERENCES

Amini, M., Abbaspour, K.C., Khademi, H., Fathianpour, N., Afyuni M. and Schulin, R. Neural network models to predict cation exchange capacity in arid regions of Iran. European J. Soil Sci., 2005; 56: 551-559.

APERI. Mahidasht-Sanjabi plain study: (phase 1) volume 2, 3, 4 and 5:climate, topography, soil and soil study. TAM consulting engineers, Ministry of Agriculture, Iran. 2004; 325 P.

Attaeian, B. Estimation of aboveground biomass carbon sequestration potential in the rangeland ecosystems of Iran. ECOPERSIA, 2016; 4 (1): 1283-1294.

Beers, T.W., Dress, P.E. and Wensel, L.C. Aspect transformation in site productivity research. J. Forestry. 1966; 64: 691–692.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. Classification and regression trees. Wadsworth International Group, Belmont. 1984; 255 P.

Boellstorff, D.L. Estimated soil organic carbon change due to agricultural land management modifications in a semiarid cereal-growing region in Central Spain, J. Arid Environ., 2009; 73: 389-392.

Demuth, H., Beale, M., Hagan, M. Neural network toolbox 6: user’s guide. Version 6.0. 2. MathWorks. Inc; 2009; 452 P.

Evrendilek, F., Celik, I., Kilic, S. Changes in soil organic carbon and other physical soil properties along adjacent Mediterranean forest, grassland, and cropland ecosystems in Turkey, J. Arid Environ., 2004; 73: 389-392.

Han X., Atsushi, T., Mitsuru, T. Effects of land cover type and topography on soil organic carbon storage in northern loess plateau, china. Plant Soil Sci. first published on: 10july 2009.

Heshmati M., Gheitury M., Parvizi Y. and Hosseini M. 2015. Effect of Converting Forest to Rainfed Lands on Spatial Variability of Soil Chemical Properties in the Zagros Forest, Western Iran. ECOPERSIA, 2015; 3(4): 1161-1174.

Krenker, A., Bester, J. and Kos, A. Introduction to the artificial neural networks. Artificial neural networks: methodological advances and biomedical applications InTech, Rijeka ISBN 2011; 978-953.

Lal, R. The role of soil organic matter in the global carbon cycle. Soil and Environ. Pollut., 2008; 116: 353-362.

McBratney, A.B., Minasny, B., Cattle, S.R. and Vervoot, R.W. From pedotransfer function to soil inference system. Geoderma. 2002; 109: 41-73.

McCullagh, J. A modular neural network architecture for rainfall estimation, Artificial Intelligence and Applications, Innsbruck, Austria. 2005; 767-772.

Nelson, D.W., Sommers, L.E., Sparks, D., Page, A., Helmke, P., Loeppert, R., Soltanpour, P., Tabatabai, M., Johnston, C. and Sumner, M. Methods of soil analysis Part 3-chemical methods, 1996; 1010.

Park, S.J. and Vlek, P.L.G. Environmental correlation of three dimensional soil spatial variability: A comparison of three adaptive. Geoderma. 2002; 109:117-140.

Sarmadian, F., Taghizadeh, R. and Akbarzadeh, A. Modeling of some soil properties using artificial neural network and multivariate regression in Gorgan province, north of Iran. Aust. J. Basic Appl. Sci., 2009; 3(1): 323-329.

Sharma, U.C, Datta, M. and Sharma, V. Soil Fertility, erosion, runoff and crop productivity affected by different farming systems. Ecopersia, 2014; 2 (3): 629-650.

Somaratne, S., Seneviratne, G. and Coomaraswamy, U. Prediction of Soil Organic Carbon across Different Land-use Patterns:A Neural Network Approach, Soil Sci. Soc. of Am. J. 2005; 69: 1580-1589.

Sparling, G.P., Wheeler, D., Wesely, E.T. and Schipper, L.A. What is soil organic matter worth?, J. Environ. Qual. 2006; 35: 548-557.

Spencer, M.J., Whitfort, T. and McCullagh, J. 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; 275 P.

Tan, Z. and Lal, R. Carbon sequestration potential estimates with changes in land use and tillage practice in Ohio, USA. Agric., Ecosys. Environ., 2005; 126: 113-121.

Wilson, J.P. and Gallant, J.C. Terrain analysis: principles and applications: John Wiley and Sons; 2000; 1208.

Zhang, B., Zhang, Y., Chen, D. and White, R.E., Li, Y. A quantitative evaluation system of soil productivity for intensive agriculture in China. Geoderma, 2004; 123: 319-331.

رهیافت‌های هوشمند برای تحلیل اهمیت مدیریت کاربری اراضی در مخزن کربن آلی خاکی در یک اکوسیستم نیمه خشک غرب ایران

یحیی پرویزی[۱] و مسیب حشمتی[۲]

[۱] استادیار پژوهشی، بخش تحقیقات حفاظت خاک و آبخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه، کرمانشاه، ایران

[۲] استادیار پژوهشی، بخش تحقیقات حفاظت خاک و آبخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه، کرمانشاه، ایران

تاریخ دریافت: ۱۴ مهر ۱۳۹۵ / تاریخ پذیرش: ۲۲ آذر ۱۳۹۵ / تاریخ چاپ: ۳۰ اسفند ۱۳۹۵

چکیده اثرات عوامل اقلیمی، خاکی، هندسی و مدیریتی در تخریب مخزن کربن آلی خاک و یا ظرفیت ترسیب کربن در این مخزن در حوزه مرگ (یک ناحیه نیمه خشک) ارزیابی شد. دو روش ناپارامتریک شامل الگوریتم درختان طبقه‌بندی و رگرسیون (CART) و شبکه‌های عصبی مصنوعی (ANN) پرسپترون چندلایه با الگوریتم پس‌انتشار خطا با روش پارامتری رگرسیون خطی چندمتغیره برای مدل نمودن و پیش‌بینی تغییرات کربن آلی خاک استفاده شد. نمونه‌برداری خاک به روش سیستماتیک تصادفی و با استفاده از نقشه واحد کاری که از انطباق نقشه‌های خاک، جهت و درصد شیب تهیه شده بود، انجام شد. نتایج نشان داد که مدل خطی خطای پیش‌بینی زیادی دارد. مدل CART با کلیه متغیرهای (فیزیکی و مدیریتی)، و شبکه عصبی مصنوعی با توپولوژی ۱-۲-۳۱ به ترتیب با پیشبینی ۸۱ و ۷۶ درصد تغییرات بالاترین قابلیت شبیه‌سازی را نشان دادند. ANN به کار رفته مقدار کربن آلی را اندکی بیش برآورد نمود اما این مدل بالاتری قابلیت را در ردیابی اثر متغیرهای مدیریتی در تغییرات کربن آلی خاک نشان داد. در همه متدها فاکتورهای مدیریتی کنترل کننده اصلی تغییرات ترسیب یا تخریب مخازن کربن آلی خاک در کاربری‌های مورد ارزیابی در این منطقه نیمه خشک بودند. مدل‌های هوشمند بالاترین حساسیت را به متغیرهای مدیریتی به‌ویژه اجزاء سناریوهای شخم و تناوب زراعی در اراضی زراعی و فاکتورهای مدیریت چرای دام در کاربری جنگل و مرتع نشان دادند و ۶۳ تا ۶۹ درصد تغییرپذیری ذخیره کربن الی خاک را مرتبط با این متغیرها تشخیص دادند.

کلمات کلیدی: CART، MLR، کربن آلی خاک، شبکه عصبی، محیط زیست نیمه‌خشک