Published online August 25, 2005

Prediction of Soil Organic Carbon across Different Land-use Patterns: A Neural Network Approach

S. Somaratne, G. Seneviratne,* and U. Coomaraswamy

ABSTRACT

Mathematical modeling has widely been used to predict soil organic carbon (SOC). 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. Artificial neural network (ANN) modeling approach that provides a tool to solve complex problems related to larger data sets was therefore used here to predict SOC contents across different land use patterns in a study conducted in Sri Lanka. Selection of variables was made using a priori knowledge of the relationships between the variables. Thus, soils of the sites were sampled and analyzed for organic C by internal heat of dilution (Ci) and external heat of dilution (Ce), and the results were presented as grams per kilogram (g kg^{-1}). In addition, some landscape attributes and environmental parameters of the sites were also collected. The predictive performance of ANN was compared with multi-linear regression (MLR) models. The best ANN model predicted the measured Ci content with R^2 of 0.92. However, comparison of the two types of models indicated less bias and high accuracy of the ANN compared with MLR in predicting Ci, but the reverse for Ce. In order to better predict Ce, it is recommended to use other architectures of neural networks and training algorithms for improving predictive accuracy. The predictive capability of the ANN developed with easily available climatic and terrain data are of importance in predicting SOC with minimum cost, labor, and time.

The soil system strongly influences the structure and function of ecosystems and acts as a buffer to global climatic change. Therefore, understanding of the processes in the soil is crucial in the context of the ecosystem management. The SOC is a vital component, since it plays a key role in soil fertility and in hydrology and acts as a sink or source of terrestrial C, which affects the concentration of atmospheric CO_2. Soil information is important in modeling ecological processes, vegetation dynamics, and forecasting agricultural potentials (Adams et al., 1990; Levine et al., 1996; Dixon et al., 1994).

The SOC estimates with certain degrees of uncertainties are available for regional to global scales. These estimates have been made either based on the existing soil databases or modeling techniques. Post et al. (1992) used global soil data up to 1-m depth across Holdridge life zones. The aggregation of soil data according to the ecosystem type and soil taxonomy has been compared, and the taxonomic approach appeared to be more meaningful in creating a real picture of spatial distribution of SOC. Attempts have been made to estimate global SOC using the pedon database and extrapolating them to soil units of the world soil map (Bohn, 1976, 1982; Batjes, 1996; Buringh, 1984; Kimble et al., 1990). The pedon database of the USDA Soil Conservation Service and related organizations has been used to estimate the regional distribution of organic C in the USA (Kern, 1994). However, previous studies indicated that there are uncertainties associated with such SOC estimates and often related to variations in soil map scales and series. As a whole the uncertainties associated with measuring and detecting changes in soil C pools remain high, both at individual sites and extrapolating site-level data to regional, national, or global scales (Vance, 2003). Accurate and precise approaches yet to be available for assessing the effect of management practices and land use change on the soil C for the purpose of incorporation of this important pool into future C accounting systems. The Kyoto Protocol, for instance, limits reporting of C sequestration activities to "measurable and verifiable" pools (Vance, 2003).

Mathematical modeling has been used to predict soil C evolution (Jenkinson and Rayner, 1977; Parton et al., 1988; Pastor and Post, 1985; Smith, 1982). These models possess the ability to simulate the complex processes in the formation and degradation of organic C and of describing the relationship between a numbers of soil properties that control soil C evolution. In these models empirical, stochastic, and mechanistic equations have been used to describe the simultaneous interactions of soil properties with SOC. These models rely on the available SOC data for predicting evolution in a given area, and certain models incorporate a limited number of SOC data points. In other cases, some of the models use interpolated or extrapolated SOC values. As a result, the model predicts SOC evolution poorly for a given area. Further, for a satisfactory prediction of SOC evolution, it is necessary to increase the number of SOC data points rather than interpolation or extrapolation of existing few SOC data points. The use of statistical or empirical models may also hinder the real relationships between the SOC and soil properties because strict statistical sampling designs are not generally used in soil sampling.

S. Somaratne, Dep. of Botany, The Open Univ. of Sri Lanka, Nawala, Nugegoda, Sri Lanka; G. Seneviratne, Institute of Fundamental Studies, Hantana Road, Kandy, Sri Lanka; U. Coomaraswamy, Vice Chancellor's Office, The Open Univ. of Sri Lanka, Nawala, Nugegoda, Sri Lanka. Received 21 Nov. 2003. *Corresponding author (gaminis@ifs.ac.lk).

Published in Soil Sci. Soc. Am. J. 69:1580–1589 (2005).
Pedology
doi:10.2136/sssaj2003.0293

Abbreviations: AIC, Akaike's information criterion; ANN, artificial neural network; Ce, soil organic carbon determined by application of external heat; CEC, cation exchange capacity; Ci, SOC determined by internal heat of dilution; LSD, least significant difference; MLR, multi-linear regression; MR, mean residuals; RMSE, root mean sums of squared error; RMSR, root mean square of residuals; SOC, soil organic carbon; TA, transformed aspect.

When dealing with an incomplete or imperfect database, Beinroth (1990) suggested a solution of either collecting new data or making better use of existing data. When it is always important to supplement and improve databases by collecting new data, the time and resources may not be available, and would be difficult to provide the exact measurement required for modeling SOC globally. To make better use of the soil data that have been collected to date, it is required to develop new tools for data analysis. Ideally, these tools should be able to input all of the data as a whole representation of C across the landscape and identify the complex linear and nonlinear relationships between soil properties. They should be able to observe the existing conditions, which occur at the time of sampling, and use this information to learn the relationships between SOC at the site and the other soil, landscape, and climate variables. In addition, they must be sensitive to missing data and must be rigorous enough to handle large quantities of data in to create scenarios, which may help in identifying key processes. With reference to such situations ANN could be a useful tool in prediction of SOC values in an area where SOC measurements are not available.

The ANNs are new analytical tools that are based on the models of neurological structures and processing functions in the brain (Anderson, 1996, 2001; Churchland and Sejnowski, 1993; Werbos, 1995). The neural networks have been widely used across different fields of studies in solving complex problems involving nonlinearity and uncertainty. The rapid increase in the computation power and new algorithms has made it simple to use the neural networks (De Wilde, 1996). There is an increasing tendency in the application of ANN in soil science. These applications include predictions of soil structure (Levine et al., 1996); hydraulic properties (Schaap et al., 1988; Pachepsky et al., 1996; Schaap et al., 2001; Schaap and Leij, 2000; Persson et al., 2002); pedotransfer functions (Minasny and McBratney, 2002; Nemes et al., 2003); environmental correlation of three-dimensional soil spatial variability (Park and Vlek, 2002); and pedometric use in soil survey (McBratney et al., 2002). Two studies have predicted SOC from soil parameters (Levine and Kimes, 1997), and from soil reflectance data using a NIR-spectrophotometer (Ingleby and Crowe, 2001) in temperate soils.

A typical neural network consists of inputs receptors, hidden layer(s) or neurons and output neurons (Haykin, 1994). The number of input neurons and output neurons are indicated by the structure of the data (De Wilde, 1996), and an empirical approach usually determines the number of neurons required in hidden layer(s) (Orchard and Phillips, 1991).

The objectives of the present study were to (i) train ANN to predict SOC across different land use patterns using easily available data in a tropical dataset, (ii) compare the predictive performance of the conventional MLR approach with the ANN and, (iii) investigate the use of environmental parameters in predicting SOC by ANN with minimum labor and cost-effectively.

MATERIALS AND METHODS

Site Characteristics

The landscape characteristics of the Kebithigollewa and Polonnaruwa sites (KBC, KBF, PLP, and PLF) are flat to undulating with elevations ranged from 90- to 115-m above mean sea level (msl). Soils are Reddish Brown Earths and Low Humic Gley soils (Rhodustalfs and Endoaqualfs, Soil Survey Staff, 2003) typical of the northern lowland region, with some alluvial soils occurring in the river valleys (Moorman and Panabokke, 1961; Panabokke, 1996). Reddish Brown earths are well to imperfectly drained and occupy the crest, upper- and mid-slopes of the undulating landscape. Sites at Madugoda (HGF and HGG) are located at altitudes of 540 to 750 m msl in the Kandy District, facing to the east of the central highlands of the island. The steep, dissected hilly and rolling terrains (Panabokke, 1996) characterize the topography of the area and, are located in the mid country wet agro-ecological region (Panabokke, 1996). The area consists of Immature Brown Loam soils (Dystrudepts, Soil Survey Staff, 2003), which are strongly acidic and developed from non-basic Precambrian crystalline rocks of the highland complex. The sites at Nuwara Eliya (NEF and NET) are situated at Hakgala Strict Natural Forest Reserve in the Nuwara Eliya district at an average elevation of 1800 m msl. The humid climate of the area has made the commonest soils found in the study area to become strongly weathered and heavily leached. These soils belong to the Red Yellow Podsolic soils (Rhodudults and Kandiudults, Soil Survey Staff, 2003) (Panabokke, 1996). Depending on the local relief the rock material from which the soils have been developed could be either in situ rock or slope colluviums or even both. In the profile the surface soil is brownish in color and sandy loam or sandy clay loam in texture. The subsoil color varies from red to yellow and the texture is sandy clay to sandy loam. The sites at Gilimale (GF and GR) are situated in a proposed forest reserve in the Ratnapura District of the Sabaragamuwa Province at an average elevation of 610 m msl. The forest reserve has the Red-Yellow Podzolic (Rhodudults and Kandiudults) soils typical of the south western lowland region, with some Reddish-brown earth (Rhodustalfs) and Immature Brown loam soil (Dystrudepts) in rolling, hilly, and steep slope (Moorman and Panabokke, 1961; Panabokke, 1996). The sites selected from the Sinharaja (SF and ST) are located in the southwestern lowland wet zone of the island at elevations from 300 to 1970 m msl. Soils of the area belong to the Red-Yellow Podsolic (Rhodudults and Kandiudults) soils (Moorman and Panabokke, 1961) and soil depth varies significantly.

Data Set

A total of 240 composite soil samples were collected from six natural forests and six land use patterns in three different eco-climatic regions of Sri Lanka (Table 1). A 100 by 100 m primary plot was laid in each site and was divided into 25 (5 by 5) grids. Twenty grids were randomly chosen to collect 10 soil cores from each. The soil samples were taken to a depth of 10 cm by carefully driving PVC tubes (internal diameter 50 mm) into the soil. The soil cores were bulked to form one composite sample representing each grid. Elevation of the sampling sites was measured at sampling time and complemented with topographic sheet (Survey Department, 1972). Slope and aspect of the terrain of the sampling site were measured using a clinometer and compass, respectively. The transformed aspect (TA), which aligns the index along a SW-NE axis, for the sites was calculated according to Beers et al. (1966) using the following equation:

Table 1. Land-use patterns, acronyms, geographic coordinates, and elevation of the selected soil sampling sites used in the study and means of selected inputs (variables) used in the development of neural networks for predicting soil organic carbon (SOC) measured by internal heat of dilution (Ci) and application of external heat (Ce) across different land-use patterns of Sri Lanka. Different letters in each column indicate significant mean difference at $P = 0.05$. Standard errors of the means are within parentheses.

Land use/acronym	Geographic coordination	Slope	Aspect	Elevation	Mean annual temperature	Mean annual rainfall	Ci	Ce
		degrees		m	°C	mm	g kg^{-1}	
			Natural vegetation					
Dry monsoonal forest PLF	7°50′ N–80°55′ E	15.00 (0.00) e	0.33 (0.00) f	86.0 (5.8) h	27.30 (0.00) a	1358.40 (0.00) d	3.03 (1.02) d	6.78 (1.59) f
Dry monsoonal forest KBF	8°37′ N–80°45′ E	5.00 (0.00) g	−0.29 (0.00) k	146.0 (8.21) g	27.30 (0.00) a	1358.40 (0.00) d	3.26 (0.88) d	8.29 (2.09) e
Grassland (Savannah) HGG	7°19′ N–80°52′ E	20.00 (0.00) d	0.64 (0.00) c	852.5 (2.56) d	24.30 (0.00) c	1946.70 (0.00) c	1.05 (0.30) h	5.82 (1.41) gf
Low montane forest GF	6°43′ N–80°43′ E	30.00 (0.00) b	0.57 (0.00) d	390.5 (5.83) f	27.10 (0.00) b	3740.20 (0.00) a	4.07 (0.57) c	11.7 (2.55) c
Montane forest NEF	6°55′ N–80°49′ E	35.00 (0.00) a	0.44 (0.00) e	1873.5 (7.59) a	27.30 (0.00) a	1358.40 (0.00) d	7.53 (2.10) a	14.18 (3.90) a
Sclerophyllous forest HGF	7°19′ N–80°88′ E	25.00 (0.00) c	−0.12 (0.00) j	851.3 (2.22) d	24.30 (0.00) c	1946.70 (0.00) c	2.69 (0.52) e	6.49 (0.50) gf
Wet evergreen lowland forest SF	6°24′ N–80°33′ E	20.00 (0.00) d	0.98 (0.00) a	977.5 (12.41) c	27.10 (0.00) b	3740.20 (0.00) a	4.12 (1.02) c	13.31 (2.63) ba
			Crops and plantation					
Chena KBC	8°37′ N–80°45′ E	5.00 (0.00) g	−0.29 (0.00) k	130.0 (0.00) g	27.30 (0.00) a	1358.40 (0.00) d	2.07 (0.61) f	5.35 (2.18) gf
Rice field PLP	7°50′ N–80°55′ E	0.00 (0.00) h	0.24 (0.00) g	70.0 (0.00) h	27.30 (0.00) a	1358.40 (0.00) d	1.33 (0.44) g	5.27 (0.90) g
Rubber GR	6°43′ N–80°43′ E	20.00 (0.00) d	0.69 (0.00) b	130.0 (16.38) g	27.10 (0.00) b	3740.20 (0.00) a	1.82 (0.54) gf	6.64 (0.87) gf
Tea Low country ST	7°19′ N–80°52′ E	10.00 (0.00) f	0.18 (0.00) h	760.0 (0.00) e	27.10 (0.00) b	3740.20 (0.00) a	2.12 (0.62) ef	11.16 (1.89) d
Tea Upcountry NET	6°55′ N–80°49′ E	35.00 (0.00) g	0.15f (0.00) i	1835.8 (17.49) b	15.40 (0.00) d	2044.00 (0.00) b	6.09 (1.16) b	12.79 (3.00) bc

$$TA = \cos(45 - \text{Aspect}) \quad [1]$$

Multiplying TA by sine value of slope angle was also performed to incorporate the effect of slope on direct-beam radiation. The mean annual rainfall and mean annual temperatures were obtained from the nearest meteorological stations of each sampling site.

The air-dried soil samples with gravel were ground and sieved to pass through 2-mm sieve, and used for the determination of SOC. Organic C was determined using a colorimetric method (Baker, 1976). Two methods of SOC oxidation (Walkley and Black, 1934) were used in determining different forms of organic C, which are easily oxidizable and inert forms in soil; heat-of-dilution of H_2SO_4 (120°C), which recovers 60 to 86% of SOC (Ci), and the application of external heat (150°C), as in the Schollenberger method (Schollenberger, 1927; Schollenberger, 1945) that produces a higher recovery of organic C and less variation in percentage of recovery among different groups of samples (Ce). Organic C determined by the Ci consists of a higher fraction of labile C than that determined by the Ce, which better reflects total C. The results of the both methods were corrected to oven dried (105°C for constant weight) soil basis and expressed as grams per kilogram (g kg^{-1}). Mean comparisons of selected input and output variables were performed with least significant difference (LSD) to assess their variability across the different land use patterns.

Development of the Neural Networks

A typical ANN consists of interconnected processing elements that are arranged in layers: an input layer, one or more hidden layers, and an output layer (Fig. 1). 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. Each layer of the ANN is connected by weights that have to be determined through a learning algorithm. 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. In a network, the individual processing elements of the network, x_i ($i = 1, 2, 3, \ldots N$) are linked through weighted connections (w_{ji}). Weighted inputs (i.e., $w_{ji}x_i$) are summed and a bias value (b_j) is added to produce a single activation level for the processing elements (I_j), input to the hidden unit, j ($j = 1, 2, 3, \ldots N$), thus:

$$I_j = \sum_{i=1}^{N} w_i x_i + b_{j0} \quad [2]$$

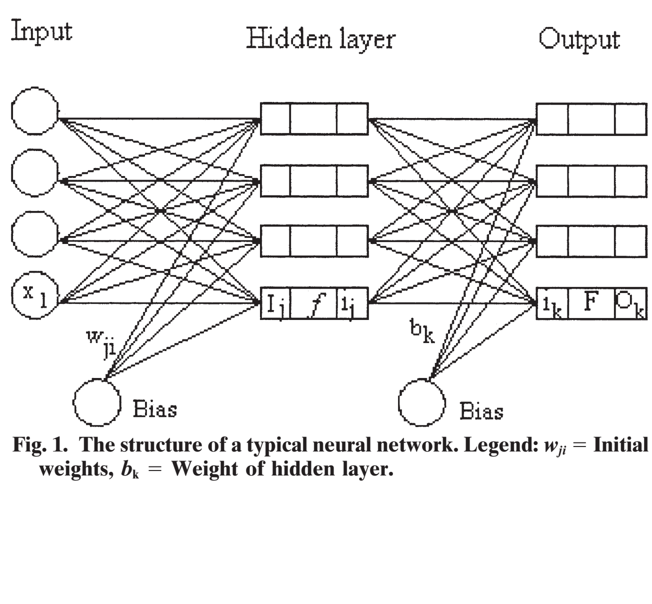

Fig. 1. The structure of a typical neural network. Legend: w_{ji} = Initial weights, b_k = Weight of hidden layer.

There are number of transfer functions available to accommodate the nonlinearity of the input-output relationship of a network (Demuth and Beale, 2000). However, the common function that has been widely used in the networks is the sigmoid or hyperbolic tangent function. Application of tansigmoid activation function to I_j produces the inputs to the hidden units (i_j).

$$i_j = f(I_j) = \tanh(I_j) = 1 - \left[\frac{2}{1 + \exp(2I_j)}\right] \qquad [3]$$

Therefore, output of the input units serves as an input for the hidden layer, that is, output of the hidden layer is computed by applying the activation function to the input i_j.

The outputs ($\boldsymbol{i}_j$) of the hidden units (L) pass to the output unit k ($k = 1, 2, 3,...N_0$) through another transfer function. The outputs of the hidden units could be indicated as:

$$i_k = \sum_{j=1}^{L} w_{kj} i_j + b_{k0} \qquad [4]$$

and the output (O_k) of the output units of the hidden layer is in the following form:

$$O_k = f(i_k) \qquad [5]$$

The weights are the adjustable parameters of the network and determined from a set of data through the process of training. The completion of training is accomplished by another process called optimization. Its objective is to minimize the sums of squared errors between the actual and predicted outputs. The desired relationship is "learned" by repeatedly presenting examples of the desired input–output relationship to the network and adjusting the model coefficients (i.e., the weights) to get the best possible agreement between the observed values and those predicted by the model. This training technique is, in general, referred to as the supervised training.

The training process is involved the following basic steps:

1. Assignment of small arbitrary values for connection weights.
2. Presentation of a training sample to the network and producing a network output.
3. Adjustment of the connection weights (w) according to the Levenberg–Marquardt algorithm (Marquardt, 1963), a combination of Newtonian gradient descent algorithm to minimize the objective function as follows:

$$w_{ij}(k + 1) = w_{ij}(k) - \eta(\partial E_p/\partial w_{ij}) + \mu[w_{ij}(k) - w_{ij}(k - 1)] \qquad [6]$$

   where η and μ are coefficients known as learning rate and momentum term respectively, and k is epoch counter.
4. Calculations of the global error function (E_p) and root mean sums of square error (RMSE) are as given below:

$$E_p = \frac{1}{2}\sum_{k=1}^{L}(y_k - O_k)^2 \qquad [7]$$

$$\text{RMSE} = \sqrt{\frac{1}{2}\sum_{k=1}^{L}(y_k - O_k)^2} \qquad [8]$$

   where y_k is the predicted value of the network. The weight is updated after the presentation of each training sample or after a number of training samples has been presented to the network. The number of training samples presented to the network between weight updates is called the epoch size (ε). Steps 2 to 4 are repeated until certain stopping criteria are met. For instance, training may be stopped when a fixed number of training samples have been presented to the network, when the E_p is sufficiently small or when there is no further improvement in the prediction obtained using an independent data set.

Determination of Network Architecture

It has been suggested by Hetch-Nielsen (1987) that the following upper limit for the number of hidden units in the hidden layer to ensure that the ANN is able to approximate any continuous function.

$$N^h \leq 2N^i + 1 \qquad [9]$$

where N^h is the number of hidden layer units and N^i is the number of inputs.

However, to ensure that the networks do not over fit the training data, the following relationship should be satisfied (Rogers and Dowla, 1994).

$$N^h \leq N^{tr}/(N^i + 1) \qquad [10]$$

where N^{tr} is the number of training samples.

Consequently, the number of hidden layer units was chosen to be the smaller of the values for N^h obtained using Eq. [9] and [10]. The structure of the network constructed (Fig. 1) consisted four input units in the input layer, four units in hidden layer, and two units in the output layer.

The Akaike information criterion (AIC, Akaike, 1973) is a performance measure that facilitates the determination of the best network size with maximum parsimony and goodness-of-fit for a particular problem. The AIC measure appreciates a network with low mean sums of squared error (MSE) but punishes networks with a large number of weights. For problems with a limited amount of data, AIC measurement can be used to determine the network that is least prone to overfit the data. Thus, it may be possible to avoid using the validation set and instead use all of the available data for training. Therefore, goodness-of-fit and parsimony of the network were assessed by calculating AIC, which was estimated according to Webster and McBratney (1989).

$$\text{AIC} = N\ln\left[\sum_{k=1}^{N}(y_k - O_k)^2\right] + 2p \qquad [11]$$

where N is the total number of data in the training set. The p is the number of weights used in the network:

$$p = (I + 1) \times H + (H + 1) \times O \qquad [12]$$

where I is the number of input units, H is the hidden units, and O is the number of output units and 1 is due to bias.

The lower values of AIC indicate the reasonable goodness-of-fit and the parsimony of the network. Considering the performance and the simplicity of the network architecture, the best network was selected and tested for its performance and accuracy in predictions of SOC.

Selection of Inputs

Selection of model inputs plays an important role in developing predictive models (Faraway and Chatfield, 1998; Kaastra and Boyd, 1995). The importance of selection of appropriate model inputs has been indicated by the differences in the predictability between ANN and conventional statistical methods. The ANNs are based on the data driven approaches whereas the statistical methods are based on model driven procedures (Chakraborty et al., 1992). In the conventional statistical methods, the structure of the model has to be predetermined using empirical or analytical approaches before the unknown model parameters could be estimated. However, the data driven approaches have the ability to determine which model inputs are critical and there is no need for a priori knowledge on the variables. However, presenting large number of inputs to the ANN models and relying on the network to determine the critical inputs often increase the network size. This may lead to decrease processing speed and increase the amount of data required to estimate the weights efficiently (Lachtermacher and Fuller, 1994).

In an ecological point of view, the abiotic and biotic components, and disturbance regime interact each other in formation of patterns in landscapes. The landform heterogeneity contributes much to the variation in climate as expected over the landscape-scale in a terrain. Each of the main components of climate such as temperature, radiation, and precipitation varies with landforms generating landscape-scale heterogeneity in the physical aspect of the terrain. The temperature varies with the radiation loading of a particular terrain, and the slope and aspect of the terrain play an important role in the variations of such parameters in a complex terrain. Further, temperature varies with the latitude and elevation of the terrain. The precipitation varies at continental scale in response to the interactions of prevailing winds and orography of a terrain. In a mountainous landscape, the orographic effect may exert a profound influence in precipitation patterns even over a few kilometers. The influence of the slope aspect has an important effect on the characteristics of the vegetation (Kutiel, 1992; Sternberg and Shosany, 2001). The vegetation (above- and belowground biomass) plays an important role as a sole input of SOC, thus any changes in the vegetation are obviously reflected in the quality and the quantity of SOC. Therefore, climatic parameters that potentially influence the characteristics and the distribution of vegetation, determine the amount of SOC in a given region or an area. The terrain attributes except slope angle have been shown to exhibit spatial dependence within the scale of sampling interval of 200 to 1000 m (Hongqing et al., 2002). In this study, the spatially structured variance (the variance due to the location of sampling sites) accounted for a large proportion of the sample variance for elevation (99%), bulk density (90%), SOC (68%), aspect (56%), and soil moisture (44%). Further, the ranges of spatial dependence (the distances within which parameters are spatially dependent) for aspect, SOC, elevation, soil moisture, and bulk density were 9810, 3070, 1120, 930, and 430 m, respectively. Most of the ecological processes depend on the physical variables, which are logistically difficult to measure. The soil variables are especially problematic over large scales, at least at the level of spatial resolution. Considering these facts selection of inputs in the present study was made with a priori knowledge of the correlations of climatic parameters and the SOC contents.

In addition, correlations between the input and output variables were also accessed and care was exercised to select the variables that easily measurable or readily available for the selected sites. On this basis, four variables such as elevation, slope with aspect of the terrain, mean annual temperature, and mean annual rainfall were selected to serve as inputs to the ANN and MLR models.

Preprocessing Data and Training Networks

Preprocessing of the data for the neural network analysis was performed in such a way that the efficiency in network training procedure is improved. In this process, the data set was standardized to have a zero mean and the standard deviation of one. Subsequently, input dataset was fully randomized before presenting to the network and subsets of data records were extracted from the total dataset for training (120), validation (60), and test (60).

A number of networks were developed for inputs variables using built-in function of neural network toolbox of MATLAB (Demuth and Beale, 1998, 2000). The best performance of network is often determined by constructing a layered network using Levenberg–Marquardt algorithm (*trainlm*). The hyperbolic tangent-sigmoid transfer function (*tansig*) and linear transfer function were used in the hidden and output layers, respectively (Hagen et al., 1996). The *trainlm* was selected and other network architecture variables were set at default values (Hagen et al., 1996), unless the network performance was satisfactory.

Adjustment of the weight was repeated, attempting to minimize the MSE between the model and the set of known outputs. The training set consisting of the inputs and known outputs, was repeatedly presented to the network. An error associated with each weight update was checked with a numerical optimization technique that is used to minimize the MSE for the training set, which is termed as epoch. The MSE for the training set approaches zero after sufficient training but at that stage the network will have lost generality and has become a poor predictor for data not used in the training. To prevent such training, network performance is tested at each epoch by computing MSE for a separate set of inputs and known outputs, which is called the validation set, and is not used to adjust the weights in the network. At certain points in the training, the MSE for validation set begins to increase, which indicates a loss of generality, therefore, training is normally stopped when the MSE for the validation set is minimized (Demuth and Beale, 1998, 2000). Initially, the experimental trials were set to 1000 epochs. The MSE used as the goal was 0.01. Training process was stopped if the goal was met before 1000 epochs. The MSE or RMSE were calculated from the differences between the network output and the target using Eq. [7] and [8], and were also used as indicators of performance of the network.

Comparing the simplicity and parsimony of network architecture, training, and validation MSE values, and the best network were selected and subjected to a number of trials with the test data set. The performance of the best architecture was judged by relationship between the output and the target that is by a low final MSE. In addition, an assessment of the performance of the neural network was made using a test data set and a regression analysis of output of the network, and the predicted and the observed SOC.

The trained ANN model was used to predict the SOC contents of the 60 soil samples (test dataset) of the different land use patterns. The accuracy of the prediction of a network is measured by the mean residual (MR), which quantifies the systematic errors between measured and predicted values. A negative value of MR indicates an underestimation and a positive value indicates overestimation.

The root of the mean square of residual (RMSR), between the observed and predicted values, and the coefficient of determination or goodness-of-fit (R^2) were also calculated. The RMSR represents the bias of the network indicating the tendency of under- or overestimation. The RMSE calculates mean accuracy, the probable magnitude of error.

The limitations of the ANN include large number of neurons requiring a significant computer memory, long training times, and convergence to local minima (Anderson, 1996). Techniques to overcome these problems include the use of a momentum factor, an adaptive learning rate and refined algorithms. The Levenberg–Marquardt method significantly reduces training time, although it demands a large computer memory (Demuth and Beale, 1998, 2000).

Multilinear Regression

For the purpose of comparison of ANN with conventional MLR, two equations were constructed one for each SOC type, Ci and Ce. The MLR analyses were performed on the training set that was already used to develop the neural networks. Two sets of coefficients were estimated using maximum R square model selection option in PROC REG of SAS (SAS, 1989). A validation dataset was used to validate the two MLR equations whereas test dataset was used to test the performances of the MLR equations. For the purpose of evaluation of the accuracy of the prediction of MLR models, MR and RMSR were calculated. In addition, R^2 between the measured and estimated values (SAS, 1989), and the AIC values for each model were also calculated. An analysis of variance was carried out on the mean sums of square errors (MSE) of ANN and MLR, and F-test at the probability level of 0.05 was employed.

RESULTS AND DISCUSSION

Variations of the selected variables across different land use patterns are given in Table 1. The variables of the sites greatly varied among the land-use patterns (e.g. slope-aspect and elevation from −0.29 to 0.98 and from 70 to 1874 m, respectively), and reflected the heterogeneity of the topography of the sites included in the study. The mean annual temperature and mean annual rainfall of the study sites ranged from 15.4 to 27.3°C, and 1358 to 3740 mm, respectively, indicating the climatological variation within the sampling sites. The SOC content as evaluated by Ci varied from 1.05 g kg^{-1} for grassland soils of the abandoned tea plantations to 7.53 g kg^{-1} in the montane forest soils. The SOC contents as evaluated by Ce varied from 5.27 g kg^{-1} in soils of the rice field in the dry zone to 14.18 g kg^{-1} in the montane forest soils. Higher SOC contents were limited to the higher elevations and it is simply due to slow microbial decomposition of soil organic matter with relatively low temperatures at higher elevations than that of the lower elevations. The SOC contents are often lower for Ci than Ce. Increased organic matter oxidation by additional heat supplied in the external-heat method can explain this difference. The analysis of variance of the MLR models of Ci and Ce indicated that the models are highly significant ($P < 0.01$). The models of Ci and Ce explained 73 and 82%, respectively of the variations of the observed C contents of the different land use patterns (Fig. 2).

The p-values of the parameters included in the MLR model of Ci indicated that all the parameters, including intercept are highly significant ($P < 0.01$; Table 2). The Ce model indicated similar significant levels except for elevation ($P > 0.05$). The residual plots of Ci and Ce are shown in Fig. 3. They reveal that there is a random distribution of residuals. However, Ci indicated a trend in which an increase in the magnitude of residues with increasing predicted SOC levels. In this regard, quadratic models are important in establishing the variation in the predicted values. The comparison of models indicates that both models could be reasonably used to model the SOC in the soils of the different land use patterns.

The relationship between the epoch and the calibration (training) and validation and test MSE values during the training of network is shown in Fig. 4. The training was stopped after 37 epochs because the validation error increased. The figure indicates that the errors associated with validation and test set are similar in characteristics. It seems that there is no significant overfitting during the training process. The plot of the predicted and observed Ci is shown in Fig. 5A. There is a

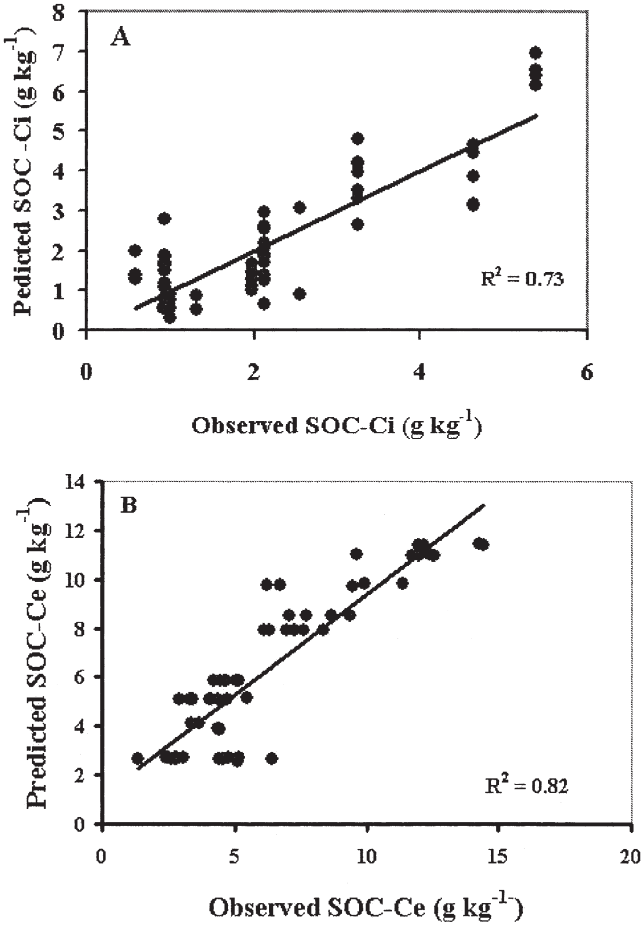

Fig. 2. Relationships between the observed and MLR model predicted soil organic carbon (SOC). The SOC measured by (A) internal heat of dilution (Ci) and (B) application of external heat (Ce) across different land use patterns of Sri Lanka.

Table 2. Estimated parameters of multi-linear regression (MLR) models developed from training data set of soil organic C measured by internal heat of dilution (Ci) and application of external heat (Ce).

Variable	Parameter estimate	Standard error	F value	$P > F$
		Ci		
Intercept	14.48603	1.53212	89.39	0.0001
Aspect	2.25259	0.18327	151.08	0.0001
Elevation	−0.00135	0.00039	11.96	0.0008
Temperature	−0.47822	0.05576	73.57	0.0001
Rainfall	0.00027	0.00000	7.91	0.0058
		Ce		
Intercept	12.70144	2.545159	24.9	0.0001
Aspect	3.98194	0.304440	171.07	0.0001
Elevation	0.00041	0.000647	0.39	0.5324
Temperature	−0.35822	0.092621	14.96	0.0002
Rainfall	0.00095	0.000158	36.07	0.0001

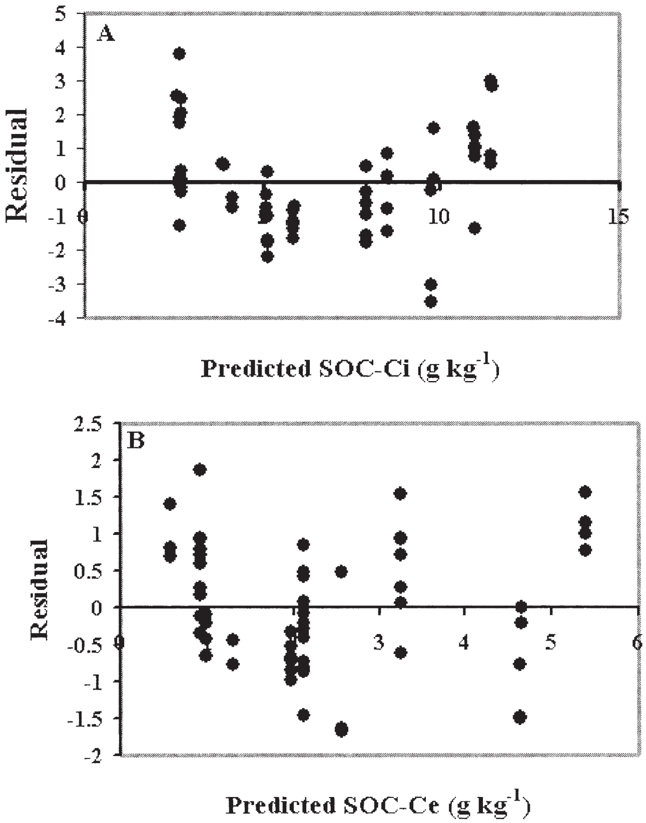

Fig. 3. Plot of residuals of multi-linear regression (MLR) model predicted soil organic carbon (SOC) as measured by (A) internal heat of dilution (Ci) and (B) application of external heat (Ce) across different land use patterns of Sri Lanka.

considerable agreement between the values with a R^2 of 92%. The relationship between ANN predicted and observed Ce is shown in Fig. 5B. The R^2 value for the predicted and observed Ce was 83%. Comparison of the R^2 values indicates that the ANN model is better in predicting Ci than Ce in the land-use patterns studied. Further, use of other network architectures with more hidden layers and different stopping criteria for network training is worthy to try out for better prediction of Ce in the land use patterns of Sri Lanka.

The AIC values of the MLR model developed for predicting Ci and Ce were −35 and 87, respectively (Table 3). The comparison of AIC values reveals that the MLR model is better in predicting Ce than Ci. This indicates that the MLR model is more suitable in predicting total C than labile C. The labile or active soil organic matter fractions are generally sensitive to the environmental parameters (Franzluebbers et al., 2001). Further, composition of microbial population that is responsible for decomposition of active soil organic matter also varies with the change in environmental parameters, making it a complex system with nonlinear relationships that are difficult to model using existing methods. The AIC values of the ANN models of Ci and Ce were −63 and −48, respectively. These values indicate that the neural networks perform better than the MLR models in predicting both Ci and Ce.

The analysis of variance of predicted values by MLR

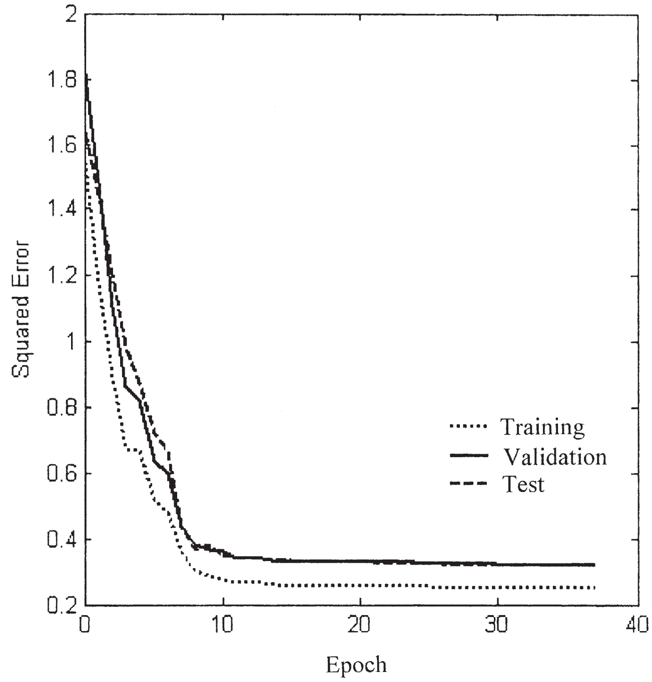

Fig. 4. Plot of training, validation and test errors during the training process of the network constructed for the prediction of soil organic C measured by internal heat of dilution (Ci) across different land-use patterns of Sri Lanka.

and ANN models indicated that there was a marginal difference between the MLR and ANN model outputs for Ci ($F = 3.777$; $P = 0.054$). However, the difference between MLR and ANN model outputs for Ce was highly significant ($F = 9.851$; $P < 0.01$).

The MR and RMSR values for the MLR and ANN models are shown in Table 3. The MR values of MLR models indicate that they underestimate Ci and overestimate Ce, representing a considerable tendency for better prediction of Ce. Their RMSR values indicate mean inaccuracies and bias in predicting the SOC. On the other hand, MR values of the ANN are lower than that of MLR. As such, comparison of MLR and ANN models reveals that MLR models are more prone to underestimate the SOC contents. The RMSR values of ANN and MLR models for Ce are 1.01062 and 0.000213, respectively, which indicates that the ANN is poor in predicting Ce, compared with MLR. Considering the overall performances, the network that consists of four variables that is, slope-aspect, elevation, mean annual temperature, and mean annual rain fall showed best performance with four-input units and one hidden layer consisting four units (Fig. 1).

Using soil parameters such as percentage of N, cation-exchange capacity (CEC)/clay ratio, and depth, Levin and Kimes (1997) showed that they are important in predicting SOC in Mollisols with a reasonable accuracy. The R^2 value obtained by them for predictive accuracy of the network that used percentage of N and CEC/clay ratio was 0.89. The number of samples included in their study was 2100 whereas that in the present study was only 240. The measures on simplicity and parsimony of the model(s) they used are not available in their report

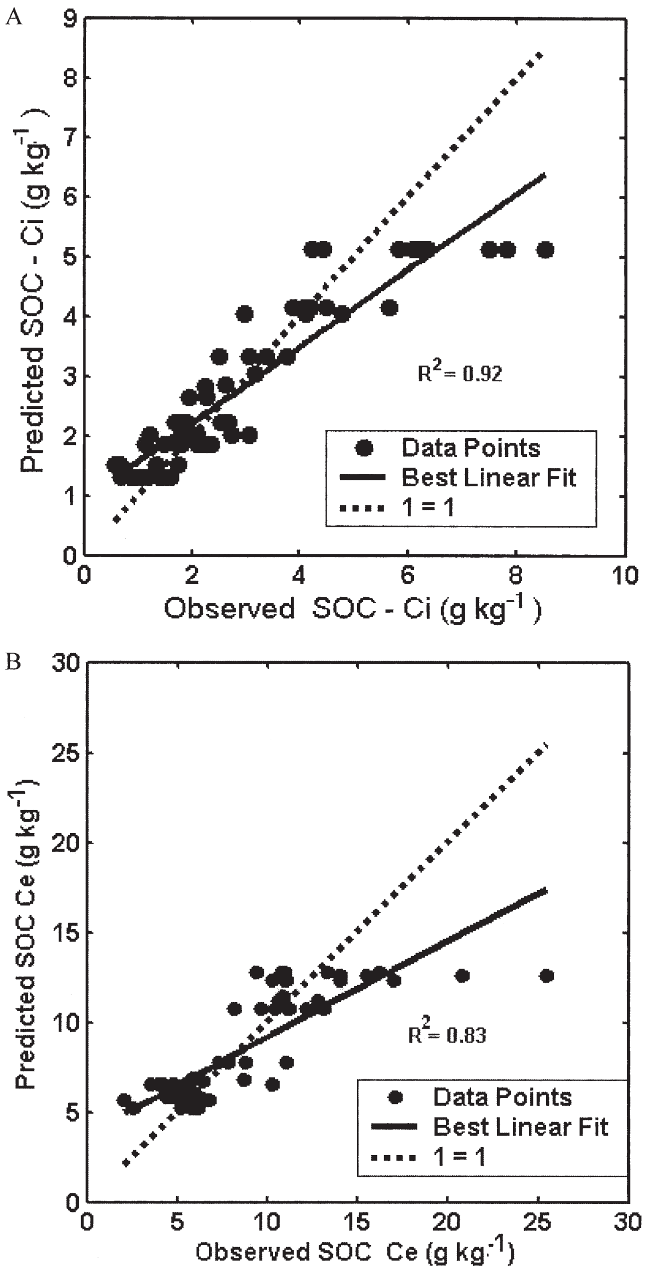

Fig. 5. Relationships between the observed and ANN model predicted soil organic carbon (SOC). The SOC measured by (A) internal heat of dilution (Ci) and (B) application of external heat (Ce) across different land use patterns of Sri Lanka.

and that prevents a direct comparison of the two models for effectiveness of predictability. However, R^2 of the best model developed in the present study was 0.92, which was higher than that of Levine and Kimes (1997). That may be due to the difference in the parameters included in the network. The calculated AIC value of the ANN model developed in the present study indicated that the model predicts the SOC particularly Ci, reasonably avoiding a considerable under or overestimation.

The overall performances of the models developed indicate that ANN models are better than the MLR models in predicting SOC across the land use patterns (Fig. 2 and Fig 5). Comparison of measured and the ANN model predicted Ci and Ce values are shown in Table 4. The ANN model predicted the SOC contents better for the Gilimale forest reserve and Sinharaja forest in the wet zone, grassland in the intermediate zone and dry zone forest at Kebithigollewa than the rest of the natural ecosystems in the study. On the other hand, SOC contents predicted by the ANN model indicated that they were very close to the observed SOC contents of natural forest sites, which are less disturbed by human interference. The SOC contents of chena soils in the dry zone, rubber plantation at Gilimale in the wet zone, and soils of the upcountry tea plantation were also well-predicted by the model.

Table 3. Comparison of performances of MLR and ANN models in predicting soil organic carbon (SOC) measured by internal heat of dilution (Ci) and application of external heat (Ce) across different land use patterns of Sri Lanka.

	Model			
	MLR		ANN	
Performance criteria	Ci	Ce	Ci	Ce
MR	−1.068	−2.5316	−0.5076	2.0106
RMSR	0.00588	0.000213	0.00215	1.01062
R^2	0.73	0.82	0.92	0.83
AIC	−35	87	−63	−48

There were slight differences in the network output with the same data set in a number of the network trials. This is because the network is automatically initialized with random weights and bias values at the start of network training. Thus, slightly different results are achieved even under training networks with identical architectures on the same data. However, the available multiple training procedures of the current software often allowed multiple training sessions to be carried out within short-time periods. The ANN model predicted average SOC approximately to those of the training and test datasets. However, networks did not predict the spread of the actual observations, that is, the range of SOC in different land-use patterns. Thus, networks are good at predicting average values, rather than the variability in a data set.

Further, comprehensive studies are required to determine the optimum number of records for the training of neural networks in predicting SOC contents in the different land use patterns in the present study. Combinations of training and test datasets and training architectural parameters that indicated more robust predictive ability will provide guidelines for structuring data, and designing and implementation of networks that could be generalized for other datasets. Mean time it is worthy to develop land-use specific ANN models for better prediction of SOC of different land use patterns.

On the data structure and feature selection, Haykin (1994) is in the opinion that the number of training examples is proportional to the number of synaptic weights in the network, which is inversely proportional to the error goal. For instance, a 0.01 (1%) mean square error goal indicates the number of training examples, which should be 100 times the number of synaptic weights in the network. The results of the present study

Table 4. Comparison of observed soil organic carbon (SOC) measured by internal heat of dilution (Ci) and application of external heat (Ce) with ANN model predicted SOC. Values within parentheses are standard error of the means.

Land use	Observed SOC-Ci	Predicted SOC-Ci	Observed SOC-Ce	Predicted SOC-Ce
	g kg^{-1}			
		Natural vegetation		
Forest (GF)	4.086 (0.401)	4.061 (0.009)	11.346 (0.511)	11.725 (0.0064)
Forest (KBF)	2.854 (0.276)	2.968 (0.234)	8.510 (0.806)	7.677 (0.551)
Forest (HGF)	2.438 (0.190)	2.909 (0.115)	6.193 (0.285)	6.717 (0.235)
Forest (PLF)	2.461 (0.370)	2.745 (0.183)	6.379 (1.164)	6.264 (0.323)
Forest (SF)	4.449 (0.411)	3.868 (0.215)	12.486 (1.502)	12.73 (0.000)
Forest (NEF)	6.620 (0.923)	5.114 (0.000)	18.981 (2.594)	12.684 (0.0001)
Grassland (HGG)	1.436 (0.105)	1.363 (0.002)	5.473 (0.461)	4.995 (0.00008)
		Crops and plantations		
Chena (KBC)	2.116 (0.225)	2.228 (0.000)	4.297 (0.253)	5.934 (0.000)
Rice field (PLP)	1.571 (0.280)	1.359 (0.000)	4.949 (0.807)	4.994 (0.000)
Rubber (GR)	1.828 (0.166)	1.930 (0.000)	5.462 (0.2833)	5.705 (0.00002)
Tea Upcountry (NET)	6.112 (0.544)	5.114 (0.000)	12.693 (1.270)	12.701 (0.000)
		Tea		
Low country (ST)	2.100 (0.233)	2.011 (0.000)	10.897 (0.621)	10.731 (0.000)
Total	2.964 (0.244)	2.767 (0.168)	8.706 (0.579)	8.333 (0.407)

indicate that environmental parameters co-vary considerably among the sites selected and it was problematic for SOC prediction in the case of general feed forward architectures. As such few environmental parameters were selected from the dataset using a priori knowledge and the readiness of the availability. However, it is worthy to try out the effect of individual input variables in the model and the sensitivity of the model as stated by Lee et al. (2003). The prediction of SOC across different land-use patterns, includes computational complexity, nonlinearity, and uncertainty, all of which neural networks are theoretically capable of handling. However, further studies are required in developing a consistent and reliable approach that generalizes the task well. The selection of soil and environmental parameters, combination of features in training data, and the selection of suitable algorithms and other architectural parameters are of importance in future studies.

CONCLUSIONS

The comparison of the two types of models indicated less bias and high accuracy of the ANN compared with MLR in predicting Ci, but the reverse for Ce. Therefore, Neural networks showed improved predictive performance over the MLR models in predicting Ci. The study revealed that further improvement in the predictive accuracy of SOC for different land-use patterns is possible through the selection of inputs to maximize nonlinear model performance. Implementation of neural network modeling using trained nets for predicting SOC in different areas thus appears rewarding. However, investigation in determining optimum inputs for nonlinear network development would be of interest in the prediction of SOC contents in different land use patterns in tropics. It is worthy to develop neural networks using larger sets of area specific data to have a better simulation of SOC across different land-use patterns.

ACKNOWLEDGMENTS

The authors express their sincere thanks to Prof. K. Tennakone, Director, Institute of Fundamental Studies, Hantana Road, Kandy, Sri Lanka for providing institutional and laboratory facilities; R.C.K. Karunaratne for assisting in laboratory analysis of soil samples; A. Lal, K.S.N. Kumara, W.R.N. Aroshana, M.A. Wijesekera for field work; K.L.A. Somapala, A. Parameshwaran and P.S.K. Rambadagalla for assisting in data entry, and N.M.I.U.K. Piyasena, Elementary Computer Laboratory, The Open University of Sri Lanka for helping in data analysis.

REFERENCES

Adams, J.M., H. Faure, L. Faure-Denard, J.M. McGlade, and F.I. Woodward. 1990. Increases in terrestrial carbon storage from the last glacial maximum to the present. Nature (London) 348:711–714.

Akaike, H. 1973. Information theory and extension of maximum likelihood principle. p. 267–281. *In* B.N. Petrov and F. Csaki (ed.) Second International Symposium on Information Theory. Akademia Kiado, Budapest.

Anderson, J.A. 1996. Introduction to neural networks. The MIT Press, Cambridge, MA.

Anderson, J.A. 2001. Introduction to neural networks. Prentice-Hall of India Pvt. Ltd., New Delhi, India.

Baker, K.F. 1976. The determination of organic carbon in soil using a probe-colorimeter. Lab. Pract. 25:82–83.

Batjes, N.H. 1996. Total carbon and nitrogen in the soils of the world. Eur. J. Soil Sci. 47:151–163.

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

Beinroth, F.H. 1990. Estimating missing soil data: How to cope with an incomplete database. Agro-technology Transfer 11:10–11 IBSNAT, Dep. of Agronomy and Soils, University of Hawaii, Honolulu, HI.

Bohn, H.L. 1976. Estimate of organic carbon in world soils. Soil Sci. Soc. Am. J. 40:468–470.

Bohn, H.L. 1982. Organic carbon in world soils. Soil Sci. Soc. Am. J. 46:1118–1119.

Buringh, P. 1984. Organic carbon in soils of the world. p. 91–109. *In* G.M. Woodwell (ed.) The role of terrestrial vegetation in the global carbon cycle: Measurement by remote sensing. SCOPE, John Wiley & Sons, New York.

Chakraborty, K., K. Mehrotra, C.K. Mohan, and S. Ranka. 1992. Forecasting the behaviour of multivariate time series using neural networks. Neural Networks 5:961–970.

Churchland, P.S., and T.J. Sejnowski. 1993. The Computational brain. The MIT Press, Cambridge, MA.

De Wilde, P. 1996. Neural network models: An analysis. Springer-Verlag, Berlin.

Demuth, A.B., and M. Beale. 1998. Neural network toolbox for use with MATLAB user's guide. Ver. 3. The Math Works Inc., Natick, MA.

Demuth, A.B., and M. Beale. 2000. Neural network toolbox: For

use with MATLAB user's guide. Ver. 4. The Math Works Inc., Natick, MA.
Dixon, R.K., S. Brown, R.A. Houghton, A.M. Solomon, M. Trexler, and J. Wisniewski. 1994. Carbon pools and flux of global forest ecosystems. Science (Washington, DC) 263:185–190.
Faraway, J., and C. Chatfield. 1998. Time series forecasting with neural networks: A comparative study using the airline data. Appl. Stat. 47:231–250.
Franzluebbers, A.J., R.L. Heney, C.W. Honeycutt, M.A. Arshad, H.H. Schomberg, and F.M. Hons. 2001. Climatic influence on active fractions of soil organic matter. Soil Biol. Biochem. 33:1103–1111.
Hagen, M.T., H.B. Demuth, and M. Beale. 1996. Neural Networks design. PWS Publishing Co., Boston, MA.
Haykin, S. 1994. Neural networks: A comprehensive foundation. Prentice-Hall Inc., Upper Saddle River, NJ.
Hetch-Nielsen, R. 1987. Kolmogorov's mapping neural networks existence theorem. p. 11–14. *In* Proceedings of the first IEEE International Joint Conference on Neural Networks. San Diego, CA.
Hongqing, W., C.A.S. Hall, J.D. Cornell, and M.H.P. Hall. 2002. Spatial dependence and the relationship of soil organic carbon and soil moisture in the Luquillo Experimental Forest, Puerto Rico. Landscape Ecol. 17:671–684.
Ingleby, H.R., and T.G. Crowe. 2001. Neural networks models for predicting organic matter content in Saskatchewan soils. Can. Biosyst. Eng. 43:7.1–7.5.
Jenkinson, D.S., and J.H. Rayner. 1977. The turnover of soil organic matter in some of the Rothamsted classical experiments. Soil Sci. 123:298–305.
Kaastra, I., and M.S. Boyd. 1995. Forecasting future trading volume using neural networks. J. Future Markets 15:953–970.
Kutiel, P. 1992. Slope aspect effect on soil and vegetation in Mediterranean ecosystem. Israel J. Bot. 41:243–250.
Kern, J.S. 1994. Spatial patterns of soil organic carbon in the contiguous United States. Soil Sci. Soc. Am. J. 58:439–455.
Kimble, J.M., H. Eswaran, and T. Cook. 1990. Organic carbon on a volume basis in tropical and temperate soils. p. 248–253. *In* M. Koshimo (ed.) Transactions of the 14th International Congress of Soil Science, Kyoto, Japan. 12–18 Aug. 1990. Vol. 5 International Soil Science Society, Kyoto, Japan.
Jenkinson, D.S., and J.H. Rayner. 1977. The turnover of soil organic matter in some of the Rothamstad classical experiments. Soil Sci. 123:298–305.
Lachtermacher, G., and J.D. Fuller. 1994. Backpropagation in hydrological time series forecasting. p. 229–242. *In* K.W. Hipel et al. (ed.), Stochastic and statistical methods in hydrology and environmental engineering. Kluwer Academic, Dordrecht.
Lee, J.H.W., Y. Huang, M. Dickman, and A.W. Jayawardena. 2003. Neural network modeling of coastal algal blooms. Ecol. Modell. 159:179–201.
Levine, E.R., and D.S. Kimes. 1997. Predicting soil carbon in Mollisols using neural networks. p. 473–484. *In* R. Lal et al. (ed.) Soil process and the carbon cycle, CRC Press, Boca Raton, FL.
Levine, E.R., D.S. Kimes, and V.G. Sigilitto. 1996. Modeling soil structure using artificial neural network. Ecol. Modell. 92:101–108.
McBratney, A.B., B. Minasny, S.R. Cattle, and R.W. Vervoot. 2002. From pedotransfer function to soil inference system. Geoderma 109:41–73.
Marquardt, D. W. 1963. An algorithm for least-square estimation of non-linear parameters. J. Soc. Ind. Appl. Math. 11:431-441.
Minasny, B., and A.B. McBratney. 2002. The neuron-m method for fitting neural-network parametric pedotransfer functions. Soil Sci. Soc. Am. J. 66:352–361.
Moorman, F.R., and C.R. Panabokke. 1961. Soils of Ceylon. Trop. Agric. 117:4–65.
Nemes, A., M.G. Shaap, and J.H.M. Wösten. 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67:1093–1193.
Orchard, G.A., and W.A. Phillips. 1991. Neural computation: Beginners' guide. Lawrence Erlbaum Associates Ltd., East Sussex, UK.
Pachepsky, Ya.A., D. Timlin, and D. Varallyay. 1996. Artificial neural network to estimate soil water relations from easily measurable data. Soil Sci. Soc. Am. J. 60:727–733.
Panabokke, C.R. 1996. Soils and Agro-ecological Environment of Sri Lanka. NARESA, Maitland Place, Colombo.
Park, S.J., and P.L.G. Vlek. 2002. Environmental correlation of three-dimensional soil spatial variability: A comparison of three adaptive techniques. Geoderma 109:117–140.
Pastor, J., and W.M. Post. 1985. Development of a Linked Forest Productivity-Soil Process Model. Oak Ridge National Laboratory Environmental Sciences Division, Publication No. 2455, NTIS, Springfield, VA.
Parton, W.J., J.W.B. Stewart, and C.V. Cole. 1988. Dynamics of C, N, P and S in grassland soils: A model. Biogeochemistry 5:109–131.
Persson, M., B. Sivakumar, R. Berndtsson, O.H. Jacobsen, and P. Schjønning. 2002. Predicting the dielectric constant–water content relationship using artificial neural networks. Soil Sci. Soc. Am. J. 66:1424–1429.
Post, W.M., W.R. Emanuel, P.J. Zinke, and A.G. Strangenberger. 1992. Soil carbon pools and world life zones. Nature (London) 29: 156–159.
Rogers, L.L., and F.U. Dowla. 1994. Optimization of ground water remediation using artificial neural networks with parallel solute transport modeling. Water Resource Res. 30:457–481.
SAS. 1989. Procedures in the SAS/STAT guide for personal computers. Ver. 6 ed. SAS Institute Inc., Cary, NC.
Schaap, M.G., F.J. Leij, and M.Th. Van Genuchten. 1988. Neural network analysis of hierarchical prediction of soil hydraulic properties. Soil Sci. Soc. Am. J. 62:847–855.
Schaap, M.G., and F.L. Leij. 2000. Improved prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten model. Soil Sci. Soc. Am. J. 64:843–851.
Schaap, M.G., F.J. Lei, and M.Th. Van Genuchten. 2001. ROSETTA: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. (Amsterdam) 251: 163–176.
Schollenberger, C.J. 1927. A rapid approximate method for determining soil organic matter. Soil Sci. 24:65–68.
Schollenberger, C.J. 1945. Determination of soil organic matter. Soil Sci. 59:53–56.
Smith, O.L. 1982. Soil microbiology: A model of decomposition and nutrient cycling. CRC Press Inc. Boca Raton. FL.
Soil Survey Staff. 2003. Keys to soil taxonomy. 9th ed. United States Department of Agriculture, Natural Resources Conservation Service, Washington, DC.
Sternberg, M., and M. Shosany. 2001. Influence of slope aspect on Mediterranean woody formation: Comparison of a semiarid and an arid site in Israel. Ecol. Res. 16:335–345.
Survey Department. 1972. One inch topography sheets. Survey Department, Colombo, Sri Lanka.
Vance, E.D. 2003. Approaches and technologies for detecting changes in forest soil carbon pools: Preamble. Soil Sci. Soc. Am. J. 67:1582.
Walkley, A., and I.A. Black. 1934. An examination of Degtjareff method determining soil organic matter and proposed modification of the chromic acid titration method. Soil Sci. 37:29–37.
Werbos, P.J. 1995. Backpropagation, basic and new developments. pp. 134-139. *In* M.A. Arbib (ed.) The handbook of brain behavior and neural networks. The MIT Press, Cambridge, MA.
Webster, R., and A.B. McBratney. 1989. On the Akaike information criterion for choosing models for variograms of soil properties. J. Soil Sci. 40:493–496.