ISSN: 0976-3031

Available Online at http://www.recentscientific.com

CODEN: IJRSFP (USA)

International Journal of Recent Scientific Research
Vol. 8, Issue, 7, pp. 18120-18128, July, 2017

International Journal of Recent Scientific Research

DOI: 10.24327/IJRSR

Research Article

ASSESSMENT OF MINE SPOIL GENESIS INFLUENCING RESTORATION IN CHRONOSEQUENCE IRON MINE OVERBURDEN SPOIL USING ARTIFICIAL NEURAL NETWORK

Pasayat, M and Patel, A.K*

School of Life Sciences, Sambalpur University, At/po- Jyoti Vihar, Burla- 768019; Dist-Sambalpur, Odisha, India

DOI: http://dx.doi.org/10.24327/ijrsr.2017.0807.0455

ARTICLE INFO

Article History:

Received 05th April, 2017
Received in revised form 21st May, 2017
Accepted 06th June, 2017
Published online 28th July, 2017

Key Words:

Artificial neural network, iron mine spoil, mine spoil genesis, QSAR, restoration

ABSTRACT

Comparative assessment of soil variables in different iron mine overburden spoil and NF soil endow with valuable information about the pace and progress of mine spoil genesis, which can be implemented for improving mine spoil restoration through sustainable use of resources. About 14 mine spoil variables were selected to develop the QSAR equation based on brute-force approach and genetic function approximation for prediction of mine spoil restoration required for fresh iron mine overburden spoil to reach the soil features of the nearby NF soil. The training and test sets with statistically best fitted with r^2 = 1.0 and r^2_{LOO} = 0.996. The predictive ANN model with 14-11-1 structure was recognized as the best model illustrating the time period required for mine spoil restoration across the sites. The standard error for the proposed model was estimated to be 0.276, which can be used as indicator of the robustness of the fit and suggested that the predicted years for mine spoil restoration based on the model is reliable. The validity of the developed model was confirmed with higher calculated value of squared correlation coefficient determination (r^2 = 0.999) and lower root mean square error (RMSE = 0.194 kPa), which revealed good predictability. Hence, IB_0 shall take ~ 38.319 years to reach the soil features of nearby NF soil depending on the variability in physico-chemical properties, enzyme activities and fungal PLFA biomarkers as sensitive and reliable indicator influencing mine spoil genesis in different age series iron mine spoil over time.

INTRODUCTION

Soil plays fundamental role in sustainable land use through the shift in microbial community structure influencing overall biodiversity, biotransformation, organic matter decomposition, nutrients turnover, pollution buffering and succession through vegetation development. Soil system is prone to degradation caused by anthropogenic activities leading to rapid deterioration of physico-chemical properties and microbiological entities, which has become the major environmental concern. The nutrient deficient topsoil with heavy metal contaminants threaten ecosystem through adverse impacts on soil quality (Sheoran *et al.,* 2008; Kumar *et al.,* 2010), which possess problems for pedogenesis (Roberts *et al.,* 1988; Hearing *et al.,* 1993), revegetation (Tordoff *et al.,* 2000; Pandey and Maiti, 2008; Bahrami *et al.,* 2010; Alavi *et al.,* 2011) and restoration of iron mine spoil (Insam and Domsch, 1988; Wong, 2003; Juwarkar *et al.,* 2009; Mukhopadhyay and Maiti, 2011; Yan *et al.,* 2013; Kujur and Patel, 2013). The assessment of mine spoil genesis is pre-requisite to implement appropriate management strategies for restoration of the legacy of mine spoil.

Monitoring microbial diversity in terrestrial ecosystems is encouraged for soil quality assessment. It is also essential for the early detection of possible decline and enables the adoption of measures to reverse such decline. Mine spoil genesis can be monitored by analyzing soil variables influencing microbial community structure and their associated functioning. Periodic monitoring can be initiated with the inventory of biodiversity such as estimation of taxonomic or functional diversity and often combined with microbial activity reflected through enzyme activities and their kinetics studies. The successive amelioration of microorganisms in different age series iron mine spoil over time brings about changes through pedogenesis, which consequently promote root growth and reduce the undesirable effects of the microclimatic conditions (Juwarkar *et al.,* 2009). Therefore, the assessment of soil variables in different age series iron mine spoil not only provide better understanding of mine spoil genesis influencing ecosystem functioning but also the implementation of appropriate management strategies for mine spoil restoration

Corresponding author:* **Patel, A.K
School of Life Sciences, Sambalpur University, At/po- Jyoti Vihar, Burla- 768019; Dist-Sambalpur, Odisha, India

(Insam and Domsch, 1988;Yan *et al.,* 2013). Besides, the use of artificial neural network is essential to validate the concept.

The determination of time period required for mine spoil restoration through experimentation is extensive and time consuming. In contrast, the alternative approach is performed through empirical mathematical modeling for prediction of the expected time required for mine spoil restoration across the sites. Such assessment can be performed by the artificial neural network (ANN), which is sophisticated computational technique for universal function approximations analogous to the neuronal function of brain (Guisan and Thuiller, 2005). It has two modes of operation such as training mode and operation/testing mode. In training mode, the neurons are trained using a particular input pattern to produce the desired output pattern. In operation/testing mode, when a trained input pattern is detected at the input, the ANN will produce its associated output. ANN is effectively used for modeling, identification, prediction and control of complex processes with nonlinearities and uncertainties (Gryglewicz, 1996; Cai *et al.,* 1996). It is useful in modeling problems in which the relation between dependent and independent variables is poorly understood and has the potentiality to identify highly complex relationships from the input-output data. Further, the back-propagation algorithm is a non-linear extension of least mean square (LSM) algorithm for multi-layer perceptrons. It is successfully applied in model-free function estimation for the pattern recognition, approximation/mapping of non-linear functions and time series prediction. The neural network is usually layered, where the layers are fully interconnected to each other. The first inputs layer receive external information datasets, which are normalized within the limit values generated from the activation functions and results in better numerical precision for mathematical operations performed by the neural network. It is supported with the second hidden layer composed of neurons, which are responsible for extracting patterns associated with the internal processes being analyzed from the network. The number of neurons in each hidden layer can vary according to the complexity or interactions among the soil variables (Kim and Gilley, 2008; Erzin *et al.,* 2010). However, the final network output is produced representing the third output layer, which results from the processing performed by the neurons in the hidden layers.

The artificial neural networks (ANNs) and feed forward artificial neural networks (FANNs) have been extensively studied to represent the process models and their beneficiary uses for industrial applications (Ungar *et al.,* 1996). Besides, ANNs have been applied to various geotechnical engineering problems such as pile capacity prediction, modeling soil behavior, site characterization, earth retaining structures, design of tunnel and underground openings, liquefaction, soil permeability and hydraulic conductivity, soil compaction, soil swelling and classification of soils (Kim and Kim, 2006; Kuo *et al.,* 2009; Banu-Ikizler *et al.,* 2010; Kalinli *et al.,* 2011; Sulewska, 2011; Chik *et al.,* 2014). In addition, the ANN approach is also applied for the prediction of organic matter (Ingleby and Crowe, 2001), soil erosion (Licznar and Nearing, 2003), hydraulic conductivity (Akbulut, 2005), volumetric moisture content (Chai *et al.,* 2008) and modeling of electrical conductivity (Davood *et al.,* 2010) of coarse grained soil samples.

Unlike analytical approaches, the ANNs require no explicit mathematical equation and no limiting assumptions of normality or linearity (MathWorks, 2005). The advantages of ANN over traditional physiology-based predictive models includes (i) the involvement of intense parallel computations during the training process, (ii) the capability of fast generalization *i.e.* once the ANN is trained for a particular system, its operation is relatively faster and the unknown input patterns can be rapidly identified in the real-time environment, (iii) the estimation of non-linear relationships between the input data and desired outputs, (iv) the data processing applications such as image recognition, (v) the classification based on land use patterns and management practices, (vi) the utilities in land drainage engineering, (vii) the estimation of crop evapo-transpiration as well as yield prediction for the new set of input conditions and thereby support the use of mechanistic simulation tools by providing the initial condition values or site-specific parameters and guide parameter estimation in agricultural machinery models (Yang *et al.,* 1997; Gopal *et al.,* 1999; Carpenter *et al.,* 1999a, 1999b; Odhiambo *et al.,* 2001; Liu *et al.,* 2001; Keller *et al.,* 2001; Behrens *et al.,* 2005; Das and Basudhar, 2008; Zhao *et al.,* 2009; Banu-Ikizler *et al.,* 2010).

Considering the tropical dry deciduous forest as natural vegetation in the study site, an attempt was made in the present study to predict the time period required for fresh iron mine overburden spoil (IB_0) to reach the soil features of the nearby forest soil based on the variability in soil properties in seven different age series iron mine spoil ($IB_0 \rightarrow IB_{25}$) in chronosequence over time through mine spoil restoration using the multivariate predictive modeling technique *i.e.* artificial neuron network (ANN). This prediction model is considered to be superior compared to non-parametric statistical benchmark methods, which provide valuable information about mine spoil genesis influencing the pace and progress of mine spoil restoration. For each dataset, the ANN predictive models were developed and all the three datasets (image-scale, field-scale and lab-scale) revealed significant network performances for training, testing and validation indicating good network generalization for predicting mine spoil restoration over time.

MATERIALS AND METHODS

Study site

The present study was carried out in the Thakurani iron mining area at Noamundi (geographical location: 85° 28' 02.61" E and 22° 8' 33.93" N), maintained by M/s. Sri Padam Kumar Jain sponge mines Private Ltd. located in the revenue district of West Singhbhum, Jharkhand, India. The study site is surrounded by a number of new, old and abandoned mine of iron ore overburdens, which were classified according to the time elapsed since inception such as fresh iron mine spoil (IB_0), 2yr (IB_2), 4 yr (IB_4), 6 yr (IB_6), 8 yr (IB_8), 15 yr (IB_{15}) and 25 yr (IB_{25}) respectively within the peripheral distance of 10 km from the core mining area. Besides, the nearby forest soil (NF) was selected adjacent to the core iron mining area for comparison of the soil variables compared to different age series mine overburden spoil over time. The district experiences semi-arid climate with annual average rainfall estimated to be 1250.43 mm as compared to the state average of 1340 mm. The mean annual temperature and humidity is around 19.67°C and 20% respectively. The study site is situated away from the mean sea level of 581m altitude.

Mine spoil sampling

Sampling was done from seven different age series iron mine overburdens and the nearby native forest in accordance with general microbiological method. During sampling, each site was divided into 3 blocks and five mine spoil samples were collected randomly from 0-15cm soil depth by digging pits of (15x15x 15) cm3 size. The samples collected from each block were referred as 'sub-samples', which were thoroughly mixed to form one 'composite sample' obtained from each overburden. Similar strategies have been followed to obtain three composite samples from each site. The samples were subjected to sieving (0.2 mm mesh size) and stored at 4ºC until analyzed.

Quantitative analysis of soil variables

Textural composition of different age series iron mine overburden spoil and nearby NF soil includes the estimation of clay (< 0.002 mm), silt (0.06-0.002 mm) and sand (2.0-0.06 mm) percentage as per the method prescribed in TSBF handbook (Anderson and Ingram, 1992). The moisture content (MC) and water holding capacity (WHC) were estimated (Mishra, 1968). Soil pH (1:2.5 ratio of soil: water) was measured with digital pH meter. Organic C (OC) was estimated through titration method suggested by Walkley and Black (Mishra, 1968). Total nitrogen (TN) was determined following Kjeldahl method (Jackson, 1958) and extractable phosphorous (EP) was estimated using chlorostannous reduced molybdophosphoric blue colour method (Olsen and Sommers, 1982).

In addition, the protease activity in different age series iron mine overburden spoil was determined by spectrophotometric method (Ladd and Butler, 1972) using sodium caseinate as substrate. The urease activity was determined by spectrophotometric method (Hoffmann and Teicher, 1961) using urea as substrate. Dehydrogenase activity was estimated spectrophotometrically through the reduction of 2,3,5-triphenylotetrazolium chloride (TTC) as electron acceptor to triphenyl formazon (TPF) (Nannipieri *et al.*, 1990; Alef and Nannipieri, 1995). Phospholipid fatty acids (PLFA) profiling of seven different age series mine overburden spoil and nearby forest soil was performed through lipid extraction based on fractionation and quantification (Buyer *et al.*, 2010).

Neural-network data-mapping model development

The back-propagation neural-network model was created using Stuttgart Neural Network Simulator package [SNNS version 4.2; Institute for Parallel and Distributed High Performance Systems (IPVR) at the University of Stuttgart, Germany] and trained using physico-chemical soil variables (textural composition, MC, WHC, pH, organic C, total N and extractable P), enzyme activities (protease, urease and dehydrogenase) and PLFA markers (18:1ω9c, 18:2ω6c, 18:3ω6c) as inputs and predicted the time required for fresh iron mine spoil to reach the soil features of nearby NF soil in years as the output. Topological structure of the neural-network model consisted of 14 input neurons in the input layer and one output neuron in the output layer to match 14:1 input-output pattern of the training datasets. One hidden layer with 11 neurons was the optimal topology for the neural-network model determined by trial and error method (Fig 1).

The evaluation criterion for determining optimal topology was the best correlation value of the training dataset. The neural network model was trained in an iterative training process using the obtained training datasets.

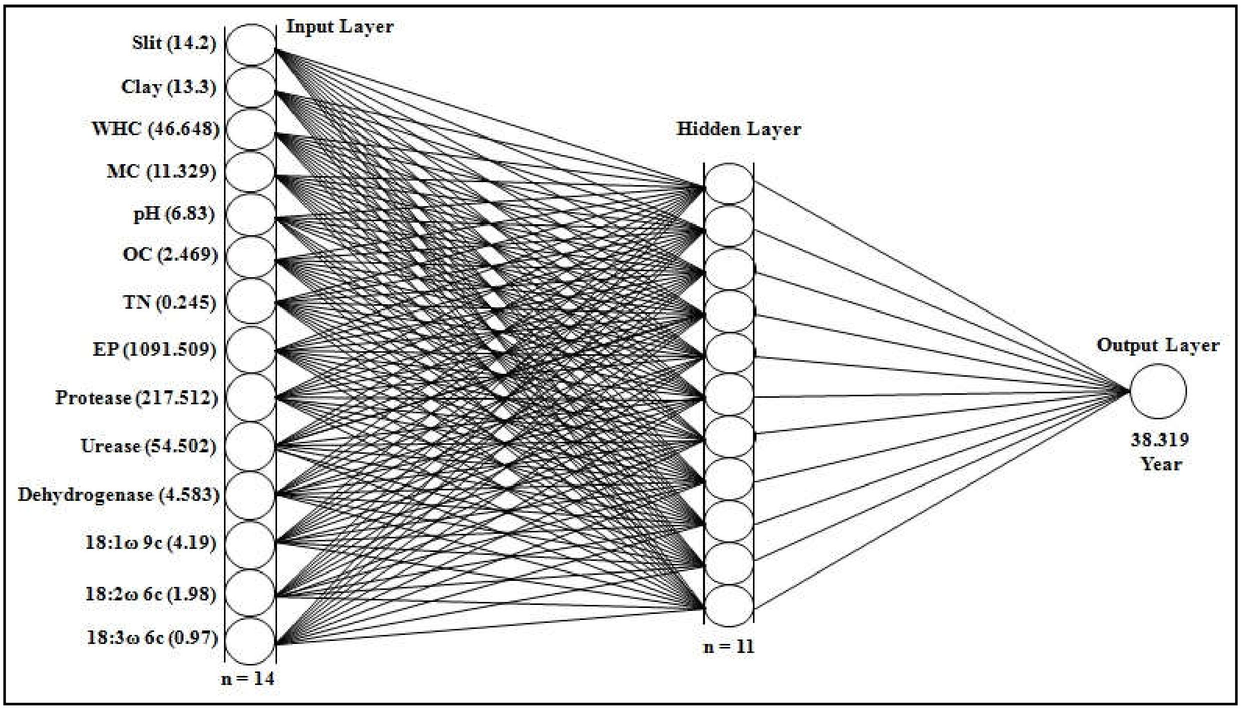


Figure 1 Layers and connection of a feed-forward back-propagating ANN. The neural network model developed here applied the sigmoid transfer function to compute the strength of interconnection between each pair of neurons.

To avoid possible bias, the order of input-output data pair in the training dataset was randomized before the training process. During training process, the back propagation training algorithm compares the estimated output value with the target value (namely the measured value), then tunes weighted values followed by connecting all the neurons to minimize the difference between the estimated and the target values until the error is smaller than the predefined level or until the number of the iteration reached a preset maximum number. The constructed model was trained with the input data for an epoch of 10,000 with 0.1 learning rate. After completion of the training process, all the weighting indices describing the interconnection strengths between the neighboring neurons are fixed and the neural network model will then be capable of mapping input variables to an estimated output promptly and accurately.

Data processing and development of prediction model

A total of 20 soil variables were used including physico-chemical parameters (silt and clay percentage, moisture content, water holding capacity, pH, organic C, total N and extractable P), enzyme activity (amylase, protease, urease and dehydrogenase), PLFAs (18:1ω9c, 18:2ω6c and 18:3ω6c), microbial CFUs (heterotrophic aerobic bacteria, sulfate reducing bacteria and actinomycetes), fungal: bacterial biomass ratio and anaerobes for the purpose. The calculated soil parameters were collected in a data matrix (D), where the rows represent mine spoil samples from seven different age series iron mine overburden spoil ($IB_0 \rightarrow IB_{25}$) and the columns represent different soil variables. In order to minimize the effect of colinearity and avoid redundancy, the correlation between different soil variables was determined and those pairs with higher relationships were considered. Among the collinear parameters, those with lowest correlation with soil properties were removed from the data matrix. Among the remaining parameters, the set of parameters that provide statistically best prediction model was selected using genetic function approximation (GFA) (Friedman, 1988) within the evolution module (ga.svl) of the MOE program.

The evolutionary genetic tool enables automated prediction modeling on the fly and is available through the SVL exchange. The GFA algorithm starts with the creation of randomly generated parameter sets. The algorithm was set up to discover the soil variables relevant for mine spoil restoration by linear polynomial terms. One hundred random initial equations with four variables were used (adding constants wherever necessary) to search for the equations of unlimited length but with the acceptable lack-of-fit (LOF) scores (Friedman, 1988), the new 'child equations' were generated using multiple linear regression method. Child equations were mutated (*i.e.* changed at "birth") 50% of the time after their generation by addition of randomly selected new terms. The number of generations of equation evolution required in the dataset was gauged by the attainment of adjusted r^2 values and minimal LOF scores. The creation of consecutive generation involves crossovers between set contents as well as mutations. Total number of crossovers was set to 50 msp 14000 with the auto-termination factor of 1000 (meaning that the calculation was stopped when the fitness function value does not change during 1000 crossovers). The equations were evaluated for statistical soundness by the Friedman LOF score, r^2, adjusted r^2, least-squares error and correlation coefficient after cross-validation statistics. The Friedman LOF score is calculated and expressed using the following equation:

LOF = LSE/ {1 – (c + dp) / m}

where; LSE is the least-square error, c is the number of basic functions in the model, d is smoothing soil variables, p is the number of soil variables, and m is the number of mine spoil samples in the training dataset.

The smoothing variables control the scoring bias between the equations of different sizes was set at default value of 1.0. The set of 14 soil variables such as physico-chemical properties (silt and clay percentage, moisture content, water holding capacity, pH, organic C, total N and extractable P), enzyme activities (protease, urease and dehydrogenase), PLFAs (18:1ω9c, 18:2ω6c and 18:3ω6c) were found to be relevant influencing mine spoil restoration in chronosequence iron mine overburden spoil over time, which were used in ANN for the development of prediction model.

Validation of prediction model

The predictive capability of the developed prediction model was further validated based on leave-one-out cross-validation. The cross-validation regression coefficient (r^2_{LOO}) was calculated based on the prediction error sum of squares (PRESS) and sum of squares of deviation of the experimental values 'Y' from their mean (SSY) using the following equation:

$$R^2_{LOO} = 1 - \frac{PRESS}{SSY} = 1 - \frac{\sum_{i=1}^{n} (Y_{\exp} - Y_{pred})^2}{\sum_{i=1}^{n} (Y_{\exp} - \bar{Y})^2}$$

where Y_{pred}, Y_{exp} and ȳ represents the predicted, observed and mean values of observed activity belonging to the training datasets of soil variables respectively. The determination coefficient in prediction using the test set (r^2 test) was calculated (Naik *et al.,* 2009, 2010).

$$R^2_{test} = 1 - \frac{\Sigma(Y_{pred_{test}} - Y_{\exp_{test}})^2}{\Sigma(Y_{\exp_{test}} - \bar{Y}_{\exp_{train}})}$$

where r_{test}^2 is the squared pearson correlation coefficient for regression calculated using Y= a + bx; a is referred to as y-intercept, b is the slope value of regression line and R_{test0}^2 is the squared correlation coefficient for regression without using y-intercept, and regression equation was Y = bx.

RESULTS AND DISCUSSION

Comparative assessment of 14 soil variables including physico-chemical properties (Pasayat and Patel, 2015), enzyme activity (Pasayat and Patel, 2016), PLFAs (Pasayat and Patel, 2017) in seven different age series iron mine overburden spoil ($IB_0 \rightarrow IB_{25}$) and the nearby forest soil (NF) have been represented (Table 1). Textural composition revealed an increasing trend in silt (7.8 - 13.5)% and clay (4.4 - 11.2)% with minimum in IB_0 and maximum in IB_{25} across the sites (Table 1). Besides, the

WHC (24.501 - 44.509)% and MC (6.643 - 10.886)% exhibited wide variation in different age series iron mine overburden spoil, which showed an increasing trend from IB_0 to IB_{25} within a span of 25 years. The mine spoil showed gradual improvement in pH from slightly acidic (6.14) in IB_0 to near neutrality (6.77) in IB_{25} in due course of time. However, the nearby NF soil exhibited relatively higher silt (14.2%), clay (13.3%), WHC (46.648%), MC (11.329%) and pH (6.83) compared to seven different age series iron mine spoil in chronosequence across the sites (Table 1). Further, the organic C (0.142 - 2.228)%, total N (0.004 - 0.187)% and extractable P (70.445 - 945.678) µg P/g spoil exhibited increasing trend from IB_0 to IB_{25} with the increase in age of iron mine overburden spoil. However, the nearby NF soil showed relatively higher organic C (2.469%), total N (0.245%) and extractable P (1091.509 µg P/g soil) compared to different age series iron mine overburden spoil across the sites (Table 1). This analysis supported by the wide variation in enzyme activity in different iron mine spoil, which was observed from IB_0 to IB_{25} *i.e.* Protease activity (2.515 - 173.755) µg tyrosine g^{-1} spoil hr^{-1}, urease activity (2.322 - 43.752) µg NH_4 g^{-1} spoil hr^{-1} and dehydrogenase activity (0.125 - 3.658) µg TPF g^{-1} spoil hr^{-1} (Table 1). The fungal PLFA biomarker also varies across the sites *i.e.* 18:1ω9c from 2.64 (IB_0) to 5.96 (IB_{25}), 18:2ω6c from 0.43 (IB_6) to 1.03 (IB_{15}) and 18:3ω6c from 0.8 (IB_{25}) to 1.34 (IB_8) respectively (Table 1).

Table 1 Comparative distribution of selected 14 parameters for ANN study of seven different age series iron mine overburden spoil ($IB_0 \rightarrow IB_{25}$) and nearby NF soil.

Parameters	Different age series mine overburden spoil from (0-15) com soil depth							NF soil
	IB_0	IB_2	IB_4	IB_6	IB_8	IB_{15}	IB_{25}	
Silt (%)	7.8	8.4	9.1	9.9	10.9	11.8	13.5	14.2
Clay (%)	4.4	5.7	6.1	6.7	7.6	8.5	11.2	13.3
WHC (%)	24.501	26.422	28.067	32.311	37.457	40.338	44.509	46.648
MC (%)	6.643	6.985	7.106	7.422	8.391	9.915	10.886	11.329
pH	6.14	6.25	6.39	6.49	6.59	6.62	6.77	6.83
Organic C (%)	0.142	0.218	0.284	0.355	0.815	1.648	2.228	2.469
Total N (%)	0.004	0.007	0.011	0.015	0.053	0.125	0.187	0.245
Extractable P (µg P/g spoil)	70.445	76.836	84.552	91.707	112.542	645.817	945.678	1091.509
Protease activity	2.515	5.163	31.253	53.753	81.3	145.754	173.755	217.51
Urease activity	2.322	4.965	7.501	9.144	12.8	28.323	43.752	54.502
DHase activity	0.125	0.313	0.542	0.708	1.56	2.813	3.658	4.583
18:1ω9c	2.64	3.62	2.67	3.14	3.11	5.96	3.25	4.19
18:2ω6c	0.63	0.91	0.43	0.78	0.78	1.03	0.88	1.98
18:3ω6c	0.83	1.02	0.97	0.86	1.34	1.14	0.8	0.97

The ANN model with these selected inputs was the best model which provide affordable results for the present study (Fig 1) based on the statistical parameters, correlation coefficient and coefficient of efficiency of the training and testing datasets (Das and Basudhar, 2008; Al-Hamed *et al.,* 2014). Several investigators suggested that the ANN developed suitable model for periodic assessment of mine spoil restoration using physico-chemical variables such as textural composition, moisture as the input variables (Das and Basudhar, 2008; Khanlari *et al.,* 2012). The results are in accordance with earlier findings, who indicated that the shift in soil quality may be due to the variation in physico-chemical properties (Zhang *et al.,* 2001; Bechmann *et al.,* 2006; Saffari *et al.,* 2009; Zhao *et al.,* 2010). Further, organic matter is reported to influence soil quality, aggregate stability and sustainability (Loveland and Webb, 2003). Thus, the dynamic changes in physico-chemical properties, organic C, available nutrients and microbial community structure need to be monitored periodically not only determine quality status but also predict the pace and progress of restoration (Sullivan *et al.,* 2005). Scientific concerns towards mine spoil restoration led to the development of computational models, connectionist system or ANN (Schaap and Leij, 1998; Stuczynski *et al.,* 1998; Koekkoek and Booltink, 1999; Pachepsky and Rawls, 1999; Wosten and Tamari, 1999; Cannon and Withfield, 2001; Patel *et al.,* 2002; Zhang and McGrath, 2004; Huang, 2009).

Further, ANNs are being widely used for predicting mine spoil restoration through the alternations in different soil variables and spatial distribution of microbial community structure (Bodaghabadi *et al.,* 2015). The practical deliverables for such study represent predictive bioclimatic model using ANN, which can be used to interpolate or extrapolate the observed interactions between the microbial taxa and their activities in response to different physico-chemical, biochemical and microbiological soil quality indices (Guisan and Harrell, 2000; Barry and Welsh, 2002; Guisan and Thuiller, 2005; Austin, 2007; Little *et al.,* 2008; Elith and Leathwick, 2009; Anadon *et al.,* 2010; Barberan *et al.,* 2012). Thus, ANNs are being widely used for predicting mine spoil genesis through the distribution and abundance of different microbial population with defined taxonomical classes in the terrestrial ecosystems (Bodaghabadi *et al.,* 2015).

Further, the values of soil vulnerability potentials to degradation in different age series mine spoil can be determined based on the variations in soil properties, which influence mine spoil restoration over time. The time period required for mine spoil restoration can be estimated through the development of prediction model based on the feed forwarded back propagation. Mine spoil samples from 21 mining sites with their efficiency towards mine spoil restoration were randomly divided into the training and test dataset of 13 and 8 mine sites respectively. Out of the total 20 parameters, 14 soil variables *i.e.* silt, clay, MC, WHC, pH, OC, TN, EP, protease, urease, dehydrogenase, 18:1ω9c, 18:2ω6c and 18:3ω6c were screened using GFA and used for the development of QSAR equation. Taking a brute-force approach, the number of variables were increases in the QSAR equation one by one and the effect of addition of a new terms with the statistical quality of the model was evaluated. The prediction model with robust prediction of the time period (in year) required for fresh iron mine spoil to reach the soil feature of NF soil has been deduced as per the following equation.

Year = - 50.8 + 4.12 Silt - 1.77 Clay - 1.87 WHC +21.8 MC - 11.3 pH - 34.0 OC + 318 TN - 0.0461 EP + 0.302 Protease + 1.56 Urease - 27.4 Dehydrogenase - 2.31 (18:1 w9c) + 3.8 (18:2 w6c) + 5.12 (18:3 w6c)

($n = 14$; $r^2 = 1.0$; LOF = 0.0001; F =1615.4; $p = 0.0001$; $r^2_{LOO} = 0.996$).

Where, n is the no. of mine spoil samples in the training set, r^2 is the squared correlation coefficient between observed and predicted years of mine spoils, *F*-test is the measure of variance that compares two models differing by one or more variables to determine if the complexity of the model correlates positively with its reliability (the model is supposed to be good if the *F*-test is above a threshold value) and r^2_{LOO} is the square of the correlation coefficient of the cross validation using the leave-one-out (loo) cross-validation technique. The prediction model is statistically best fitted (r^2= 1.0, r^2_{LOO} = 0.996) and consequently used for the prediction of years of mine spoil of training and test sets (Tables 2 and 3).

Table 2 Statistical assessment of QSAR models for the estimation of predicted year for mine spoil restoration with varying numbers of soil variables in the training set.

Sites	Observed Year	Predicted Year
IB0_S1	0.00	0.2929411
IB0_S2	0.00	-0.0109372
IB2_S1	2.00	2.5053459
IB4_S1	4.00	3.8770059
IB4_S2	4.00	3.6965352
IB6_S1	6.00	5.9890578
IB6_S2	6.00	5.8996976
IB8_S1	8.00	8.3583512
IB8_S2	8.00	8.222339
IB15_S1	15.00	15.176928
IB15_S2	15.00	15.008596
IB25_S1	25.00	24.639342
IB25_S2	25.00	24.955494

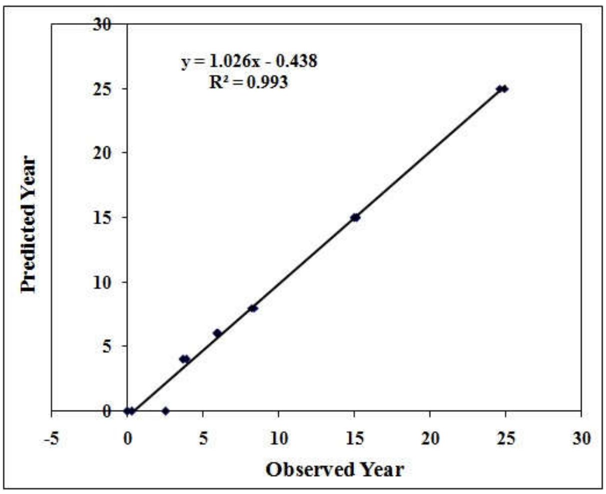
Table 3 Statistical assessment of QSAR models for the estimation of predicted year for mine spoil restoration with varying numbers of soil variables in the test set.

Sites	Observed Year	Predicted Year
IB0_S3	0.00	0.141002
IB2_S2	2.00	2.0002584
IB2_S3	2.00	2.0002716
IB4_S3	4.00	4.0002733
IB6_S3	6.00	5.9443777
IB8_S3	8.00	8.2903451
IB15_S3	15.00	15.001593
IB25_S3	25.00	24.797

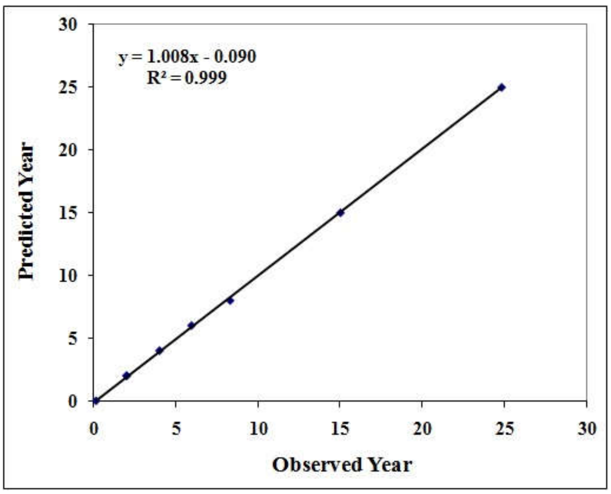
The quality of prediction models for the training set is shown (Fig 2). The r^2 and r^2_{LOO} values of the model corroborate the criteria for a highly predictive model.

The standard error for the proposed model was estimated to be 0.276, which can be used as an indicator of robustness of the fit and suggests that the predicted years of mine spoils based on the model is reliable. Similarly, the quality of prediction models for the test set (Fig 2b). The overall root mean square error (RMSE) between the observed and predicted years was found to be 0.194, which revealed good predictability. The squared correlation coefficient between the observed and the predicted years for the test set is also significant (r^2 = 0.999) (Fig 2b). The estimated correlation coefficient between observed and predicted years with intercept (r^2) and without intercept (r^2_0) is 0.9992 and 0.9991, respectively. The value of $[(r^2 - r^2_0)/r^2]$ = (0.9992 – 0.9991)/0.9992 = 0.0001 is less than the stipulated value of 0.1.

The ANN is called feed-forwarded back-propagation algorithm, which is being eligible for estimation, classification and to be useful in non-linear structural models (Demuth *et al.,* 2007). The root mean square error (RMSE) and mean error (ME) were also used to assess the accuracy of predicted model (Zhao *et al.,* 2009). Further, the better correlation between ANN predicted values and measurement from the test dataset soil parameters was observed (Fig 2). The study suggested that the ANN approach can be implemented to develop prediction model that can able to predict the pace and progress of mine spoil restoration with high efficiency and accuracy. Further, the genetic algorithm (GA) was coupled with back-propagating network analysis to optimize the informative variables and improve the accuracy of the proposed model (Karimi and Yousefi, 2012). It has wide scope of applications in data mining, where the hidden information is mined from large and distributed databases. Therefore, the application of artificial neural network with genetic algorithm (ANN-GA) facilitates more predictability and accuracy towards the assessment of time period required for mine spoil restoration across the sites. The ANN prediction model is used to determine the time required for fresh mine overburden spoil as par with the soil features of nearby NF soil taking into account the input values of 14 soil variables influencing mine spoil restoration in different age series iron mine spoil was estimated to be approximately 38.319 years.



(a)



(b)

Figure 2 Quantitative structure-activity relationship (QSAR) model revealed the relationship between the predicted and observed year for (a) training set; (b) test set soil parameters.

CONCLUSION

The quality assessment will assist to determine soil variables for the development of 'minimum datasets' (MDS) involved in determining the quality thresholds set for each biological soil quality indicator depending on the impact of anthropogenic activities or land degradation over time. The multivariate predictive modeling based on ANN-GA was designed using different soil variables to validate the network generalization to predict the time period required for mine spoil restoration. The study indicated that ANN based predictive model is considered as powerful tool in predicting the consolidation parameters more accurately. About 14 parameters were selected as input soil variables based on genetic function approximation, which influence mine spoil genesis in different age series iron mine spoil over time reflecting mine soil restoration. The proposed ANN model revealed that (i) ANN model with 14-11-1 structure was recognized as the best model for predicting the time period required for mine spoil restoration. The validity of the developed model was confirmed by higher calculated value of squared correlation coefficient determination (r^2 = 0.999) and lower root mean square error (RMSE = 0.194 kPa); (ii) the contribution analysis using input parameters on output and *vice versa* revealed that the soil variables used in ANN for the development of prediction model are highly interrelated. The study based on ANN predictive model determine the time period required for fresh iron mine overburden spoil to reach the soil features of the nearby forest soil shall take ~ 38.319 years. Therefore, the 14 soil variables can be used as the 'minimum datasets' for monitoring mine spoil genesis to determine the pace and progress of mine spoil restoration pertaining to ecosystem sustainability.

Acknowledgements

The authors are thankful to the Head, School of life Sciences, Sambalpur University for providing necessary laboratory facilities. The investigation was made possible through the support rendered by the mining authority during sampling. In particular, the authors are indebted to many, who helped in the laboratory as well as for computation of statistical analysis.

References

Akbulut, S. 2005. Artificial neural networks for predicting the hydraulic conductivity of coarse grained soils. *Eurasian Soil Science*, 38, 392-398.

Alavi, R., Alinejad-Rokny, H., and Zadeh, M.S. 2011. Prioritizing coercive plant species in Choghart iron mine desert region. *Australian Journal of Basic and Applied Sciences*, 5(12), 1075-1078.

Alef, K., and Nannipieri, P. 1995. Methods in applied soil microbiology and biochemistry (Eds). Academic Press, San Diego, London, pp. 214-218.

Al-Hamed, S.A., Wahby, M.F., and Aboukarima, A.M. 2014. Artificial neural network for soil cohesion and soil internal friction angle prediction from soil physical properties. *International Research Journal of Agricultural Science and Soil Science*, 4(5), 85-94.

Anadon, J.D., Gimenez, A., and Ballestar, R. 2010. Linking local ecological knowledge and habitat modelling to predict absolute species abundance on large scales. *Biodiversity Conservation*, 19, 1443-1454.

Anderson, J.M., and Ingram, J.S.I. 1992. Tropical soil biology and fertility. A handbook of methods. (II Eds), Oxford University Press, USA.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, 200, 1-19.

Bahrami, A., Emadodin, I., Atashi, M.R., and Bork, H.R. 2010. Land-use change and soil degradation: a case study, North of Iran. *Agriculture and Biology Journal of North America*, 4, 600-605.

Banu Ikizer, S., Aytekin, M., Vekli, M., and Kocabas, F. 2010. Prediction of swelling pressures of expensive soils using artificial neural networks. *Advances in Engineering Software*, 41, 647-655.

Barberan, A., Bates, S.T., Casamayor, E.O., and Fierer, N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. *International Society for Microbial Ecology Journal*, 6, 343-351.

Barry, S.C., and Welsh, A.H. 2002. Generalized additive modelling and zero inflated count data. *Ecological Modelling*, 157, 179-188.

Bechmann, J., Conteras, K., Hartge, K.H., and MacDonald. 2006. Comparison of soil strength data obtained in situ with penetrometer and with van shear test. *Soil and Tillage Research*, 87,112-118.

Behrens, T., Forster, H., Scholten, T., Steinrucken, U., Spies, E.D., and Goldschmitt, M. 2005. Digital soil mapping using artificial neural networks, *Journal of Plant Nutritional and Soil Sciences*, 168, 1-13.

Bodaghabadi, M.B., Casasnovas, J.A.M., Hasan Salehi, M.H., Mohammadi, J., Borujeni, I.E., Toomanian, N., and Gandomkar, A. 2015. Digital soil mapping using artificial neural networks and terrain-related attributes, *Pedosphere*, 25(4), 580-591.

Buyer, J.S., Teasdale, J.R., Roberts, D.P., Zasada, I.A., and Maul, J.E. 2010. Factors affecting soil microbial community structure in tomato cropping systems. *Soil Biology and Biochemistry*, 42, 831-841.

Cai, Z.X., Wang, Y.N., and Cai, J.F. 1996. A real-time expert control system, *Artificial Intelligence Engineering*, 10, 311-317.

Cannon, A.J., and Whitefield, P.H. 2001. Modelling transient pH depressions in coastal streams of British Columbia using neural networks. Journal of AM, *Water Resources Association*, 37, 73-89.

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., and Woodcock, C.E. 1999a. A neural network method for mixture estimation for vegetation mapping. *Remote Sensing of Environment*, 70(2), 138-152.

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., and Woodcock, C.E. and Franklin, J. (1999b). A neural network method for efficient vegetation mapping. *Remote Sensing of Environment*, 70(3), 326-338.

Chai, S.S., Veenendaal, B., West, G., and Walker, J.P. 2008. Backpropagation neural network for soil moisture retrieval using Nafe'05 data: a comparison of different training algorithms. The International Archives of the Photo-grammetry. *Remote Sensing and Spatial Information Sciences*, Beijing, 37, 1345-1350.

Chik, Z., Aljanabi, Q.A., Kasa, A., and Taha, M.R. 2014. Tenfold cross validation artificial neural network modeling of the settlement behaviour of a stone column under a highway embankment. *Arabian Journal of Geosciences,* 7(11), 4877-4887.

Das, S.K., and Basudhar, P.K. 2008. Prediction of residual friction angle of clays using artificial neural network. *Engineering Geology*, 100(3/4), 142-145.

Davood, N.K., Shorafa, M., Omid, M., and Mahmou, F.S. 2010. Application of artificial neural networks in modelling soil solution electrical conductivity. *Soil Science*, 175(9), 432-437.

Demuth, H., Beale, M., and Hagan, M. 2007. Neural network toolbox user's guide for use with matlab. Mathworks Inc, Natick.

Elith, J., and Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution and Systematics*, 40, 677-697.

Erzin, Y., Rao, B.H., Patel, A., Gumaste, S.D., and Singh, D.N. 2010. Artificial neural network models for predicting electrical resistivity of soils from thermal resistivity. *International Journal of Thermal Sciences*, 49 (1), 118-130.

Friedman, J. 1988. Multivariate adaptive regression splines, Technical Report No. 102, Laboratory for computational statistics, Department of Statistics, Stanford University.

Gopal, S., Woodcock, C.E., and Strahler, A.H. 1999. Fuzzy neural network classification of global land covers from a 1° AVHRR data set. *Remote Sensing of Environment*, 67(2), 230-243.

Gryglewicz, G. 1996. Effectiveness of high temperature pyrolysis in 475 sulfur removal from coal. *Fuel Process Technology*, 46(3), 217-26.

Guisan, A., and Harrell, F.E. 2000. Ordinal response regression models in ecology. *Journal of Vegetation Sciences*, 11, 617-626.

Guisan, A., and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecological Letter*, 8, 993-1009.

Hearing, K.C., Daniels, W.L., and Roberts, J.A. 1993. Changes in mine soil properties resulting from overburden weathering. *Journal of Environmental Quality*, 22, 194-200.

Hoffmann, G.G., and Teicher, K. 1961. Ein kolorimetrisches verfahren zur bestimmung der urease-aktivitat in boden. *Zeitschrift Fur Pflanzenernahrung Dungung Bodenkunde*, 91, 55-63.

Huang, Y. 2009. Advances in artificial neural networks-methodological development and application. *Algorithms*, 2, 973-1007.

Ingleby, H.R., and Crowe, T.G. 2001. Neural network models for predicting organic matter content in Saskatchewan soils. *Canadian Biosystems Engineering*, 43, 7-12.

Insam, H., and Domsch, K.H. 1988. Relationship between soil organic carbon and microbial biomass on chronosequences of reclamation sites. *Microbial Ecology*, 15, 177-188.

Jackson, M.L. 1958. Soil Chemical Analysis. Prentice-Hall, Englewood Cliffs, N.J., USA, pp. 485.

Juwarkar, A.A., Yadav, S.K., Thawale, P.R., Kumar, P., Singh, S.K., and Chakrabarti, T. 2009. Developmental strategies for sustainable ecosystem on mine spoil dumps: a case of study. *Environment Monitoring Assess*, 157, 471-481.

Kalinli, A., Acar, M.C., and Gunduz, Z. 2011. New approaches to determine the ultimate bearing capacity of shallow foundations based on artificial neural networks and ant colony optimization. *Engineering Geology*, 117(1), 29-38.

Karimi, H. and Yousefi, F. (2012). Application of artificial neural network-genetic algorithm (ANN-GA) to correlation of density in nano-fluids. *Fluid Phase Equilibria*, 336: 79-83.

Keller, A., Von Steiger, B., Vander Zee, S.T., and Schulin, R. 2001. A stochastic empirical model for regional heavy metal balances in agroecosystems. *Journal of Environmental Quality*, 30, 1976-1989.

Khanlari, G.R., Heidari, M., Momeni, A.A., and Abdilor, Y. 2012. Prediction of shear strength parameters of soils using artificial neural networks and multivariate regression methods. *Engineering Geology*, 131, 11-18.

Kim, M., and Gilley, J.E. 2008. Artificial neural network estimation of soil erosion and nutrient concentration in runoff from land application areas. *Computers and Electronics in Agriculture*, 64(2), 268-275.

Kim, Y.S., and Kim, B.T. 2006. Use of artificial neural networks in the prediction of liquefaction resistance of sands. *Journal of Geotechnical Geoenvironmental Engineering*, 132(11), 1502-1504.

Koekkoek, E.J.W., and Boltink, H. 1999. Neural network models to predict soil water retention. *European Journal of Soil Sciences*, 50, 489-495.

Kujur, M., and Patel, A.K. 2013. Comparative assessment of physico-chemical properties influencing microbial biomass as biomarker in monitoring soil status on chronosequences of iron mine overburden spoil. *International Journal of Environmental Sciences*, 3(5), 1656-1670.

Kumar, V., Rao, T.V., Rao, S., Prabhakar, S., and Bhaskar Raju, G. 2010. Reverse flotation studies on an Indian low grade iron ore slimes. *International Journal of Engineering Science and Technology*, 2(4), 634-645.

Kuo, Y.L., Jaksa, M.B., Lyamin, A.V., and Kaggwa, W.S. 2009. ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil. *Computer Geotechnology*, 36(3), 503-516.

Ladd, J.N., and Butler, J.H.A. 1972. Short term assay of soil proteolytic enzymes activities using proteins and dipeptide derivatives as substrates. *Soil Biology and Biochemistry*, 4, 19-30.

Licznar, P., and Nearing, M.A. 2003. Artificial neural networks of soil erosion and runoff prediction at the plot scale. *Catena*, 51, 89-114.

Little, A.E., Robinson, C.J., Peterson, S.B., Raffa, K.F., and Handelsman, J. 2008. Rules of engagement: interspecies interactions that regulate microbial communities. *Annual Review of Microbiology*, 62, 375-401.

Liu, J., Goering, C.E., and Tian, L. 2001. A neural network for setting target yields. *American Society of Agricultural and Biological Engineers*, 44, 705-713.

Loveland, P., and Webb, J. 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. *Soil Tillage Resources*, 70, 1-18.

MathWorks, 2005. Neural network toolbox. Natick, Mass. The MathWorks, Inc.

Mishra, R. 1968. Ecology Work Book. Oxford IBH, New Delhi.

Mukhopadhyay, S., and Maiti, S.K. 2011. Mine spoil reclamation due to tree plantation: a chronosequence study. *African Journal of Basic Applied Science*, 3, 210-218.

Naik, P.K., Alam, A., Malhotra, A., and Rizvi, O. 2010. Molecular modeling and structure-activity relationship of podophyllotoxin and its congeners. *Journal of Biomolecular Screening*, 15(5), 528-540.

Naik, P.K., Sindhura., Singh, T., and Singh, H. 2009. Quantitative structure - activity relationship (QSAR) for insecticides: development of predictive in vivo insecticide activity models. SAR and QSAR in *Environmental Research*, 20(5/6), 551-566.

Nannipieri, P., Grego, S., and Ceccanti, B. 1990. Ecological significance of the biological activity in soil. *Soil Biochemistry*, 6, 293-355.

Odhiambo, L.O., Yoder, R.E. Yoder, D.C., and Hines, J.W. 2001. Optimization of fuzzy evapotranspiration model through neural training with input-output examples. *Transactions of American Society of Agricultural Engineers*, 44(6), 1625-1633.

Olsen, S.R., and Sommers, L.E. 1982. Phosphorous. In: Methods of soil analysis, P-II, Miller, R.H. and Keeney, D.R. (Eds), *American Society of Agronomy*, Inc, Madison, WI.

Pachepsky, Y.A., and Rawls, W.J. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. *Soil Science Society of American Journal*, 63, 1748-1757.

Pandey, S., and Maiti, T.K. 2008. Physicochemical and biological characterization of slag disposal site at Burnpur, West Bengal. *Pollution and Research*, 27(2), 345-348.

Pasayat, M., and Patel, A.K. 2015. Assessment of physico-chemical properties influencing mine spoil genesis in chronosequence iron mine overburden spoil and implications of soil quality. *International Journal of Current Microbiology and Applied Sciences*, 4(6), 1095-1110.

Pasayat, M., and Patel, A.K. 2016. Contribution of soil physico-chemical properties influencing microbial biomass used as biomarkers for mine spoil genesis. *Research Journal of Pharmaceutical*, *Biological* and *Chemical Sciences*, 7(5), 738-747.

Patel, R.M., Prasher, S.O., Goel, P.K., and Bassi, R. 2002. Soil salinity prediction using artificial neural networks. *Journal of American Water Resources*, 38, 91-100.

Roberts, J.A., Daniels, W.L., Bell, J.C., and Burger, J.A. 1988. Early stages of mine soil genesis as affected by top soiling and organic amendments. *Soil Science Society of America Journal*, 52, 730-738.

Saffari, M., Yasrebi, J., Sarikhani, F., Gazni, R., Moazallahi, M., Fathi, H., and Emadi, M. 2009. Evaluation of artificial neural network models for prediction of spatial variability of some soil chemical properties. *Research Journal of Biological Sciences*, 4, 815-820.

Schaap, M.G., and Leji, F.J. 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil Tillage Resources*, 47, 37-42.

Sheoran, S., Sheoran, V., and Poonia, P. 2008. Rehabilitation of mine degraded land by metallophytes. *Mining Engineers Journal*, 10(3), 11-16.

Stuczynski, T., Pauly J., and Terelak, H. 1998. Neural computing approach to soil monitoring systems in Poland. *European Soil Bureau-Research Report*, 4, 321-328.

Sulewska, M.J. 2011. Applying artificial neural networks for analysis of geotechnical problems. *Computer Assist Mechanical Engineering Sciences*, 18, 231-241.

Sullivan, D.G., Shaw, J.N., Rickman, D., Mask, P.L., and Luvall, J.C. 2005. Using remote sensing data to evaluate surface soil properties in Alabama ultisols. *Soil Science*, 170, 954-968.

Tordoff, G.M., Baker, A.J.M., and Willis, A.J. 2000. Current approaches to the revegetation and reclamation of metalliferous mine wastes. *Chemosphere*, 41, 219-228.

Ungar, L.H., Hartman, E.J., Keeler, J.D., and Martin, G.D. 1996. Process modelling and control using neural networks. *American Institute of Chemical Engineering Symposium Series*, 92, 57-66.

Wong, M.H. 2003. Ecological restoration of mine degraded soils with emphasis on metal contaminated soils. *Chemosphere*, 50, 775-780.

Wosten, J.H.M., and Tamari, S. 1999. Application of artificial neural networks for developing pedotransfer functions of soil hydraulic parameters. *Geophysical Monograph*, 108, 235-241.

Yan, D., Zhao, F., and Sun, O.J. 2013. Assessment of vegetation establishment on tailing dam at an iron mining site of Suburban Beijing, China, 7 years after reclamation with contrasting site treatment methods. *Environmental Management*, pp. 1-10.

Yang, C.C., Prasher, S.O., Lacroix, R., Sreekanth, S., Patni, N.K., and Masse, L. 1997. Artificial neural network model for subsurface-drained farmlands. *Journal of Irrigation and Drainage Engineering*, 123(4), 285-292.

Zhang, B., Zhao, Q.G., Horn, R., and Baumgartl, T. 2001. Shear strength of surface soil as affected by soil bulk density and soil water content. *Soil and Tillage Research*, 59, 97-106.

Zhang, C., and McGrath, D. 2004. Geostatistical and GIS analyses on soil organic carbon concentrations in grassland of south-eastern Ireland from two different periods. *Geoderma*, 119, 261-275.

Zhao, Z., Chow, T.L., Rees, H.W., Yang, Q., Xing, Z., and Meng, F.R. 2009. Predict soil texture distributions using an artificial neural network model. *Computers and electronics in agriculture*, 65, 36-48.

Zhao, Z., Yang, Q., Benoy, G., Chow, T.L., Xing, Z., Rees, H.W., and Meng, F.R. 2010. Using artificial neural network models to produce soil organic carbon content distribution maps across landscapes. *Canadian Journal of Soil Sciences*, 90, 75-87.