Archives of Agronomy and Soil Science
Vol. 57, No. 5, August 2011, 549–565

Comparing multivariate regression and artificial neural network to predict barley production from soil characteristics in northern Iran

Shamsollah Ayoubi[a]* and Kanwar Lal Sahrawat[b]

[a]*Department of Soil Science, College of Agriculture, Isfahan University of Technology, 84156-83111 Isfahan, Islamic Republic of Iran;* [b]*Global-Theme Agroecosystems, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru (near Hyderabad), India*

(*Received 5 October 2009; final version received 17 January 2010*)

In this study artificial neural network (ANN) models were designed to predict the biomass and grain yield of barley from soil properties; and the performance of ANN models was compared with earlier tested statistical models based on multivariate regression. Barley yield data and surface soil samples (0–30 cm depth) were collected from 1 m^2 plots at 112 selected points in the arid region of northern Iran. ANN yield models gave higher coefficient of determination and lower root mean square error compared to the multivariate regression, indicating that ANN is a more powerful tool than multivariate regression. Sensitivity analysis showed that soil electrical conductivity, sodium absorption ratio, pH, total nitrogen, available phosphorus, and organic matter consistently influenced barley biomass and grain yield. A comparison of the two methods to identify the most important factors indicated that while in the ANN analysis, soil organic matter (SOM) was included among the most important factors; SOM was excluded from the most important factors in the multivariate analysis. This significant discrepancy between the two methods was apparently a consequence of the non-linear relationships of SOM with other soil properties. Overall, our results indicated that the ANN models could explain 93 and 89% of the total variability in barley biomass and grain yield, respectively. The performance of the ANN models as compared to multivariate regression has better chance for predicting yield, especially when complex non-linear relationships exist among the factors. We suggest that for further potential improvement in predicting the barley yield, factors other than the soil properties considered such as soil micronutrient status and soil and crop management practices followed during the growing season, need to be included in the models.

Keywords: artificial neural network; barley yield; multivariate regression; soil properties

Introduction

Crop productivity across fields is highly variable as a result of complex interactions among different factors such as soil properties, topography and management practices (Godwin and Miller 2003). An understanding of the effects of soil properties on the performance of the strategic cereal crops in the semiarid and arid regions

*Corresponding author. Email: ayoubi@cc.iut.ac.ir

ISSN 0365-0340 print/ISSN 1476-3567 online

DOI: 10.1080/03650341003631400
http://www.informaworld.com

provides valuable information for enhancing agricultural productivity (Ayoubi et al. 2009). Soil variability occurs within the field and landscape as a result of interactions among pedogenic factors including climate, topography, and geology as well as land use and management practices (Quine and Zhang 2002; Ayoubi et al. 2007).

The within-field variability in soil properties influences soil processes such as water and nutrient movement and their redistribution and supply to plants, root growth and sustenance; and the variability also influences the crop response to management and the susceptibility of soil to degradation (Shukla et al. 2004). Knowledge of the variability of soil properties has a vital role in selecting as well as effectively applying management decision in the field (Vieira and Paz Gonzalez 2003; Shukla et al. 2004). Determining, which soil factors to base a management decision on, is often a complex process due to the interactions among various factors that affect crop yield.

Several techniques have been applied to determine the relationships between crop yield and soil or landscape properties in an attempt to identify important factors affecting yield production. Correlation and multiple linear regressions (MLR) are commonly used for such purposes (Khakural et al. 1999; Kravchenko and Bullock 2000; Adams et al. 2004), but the results are often not satisfactory. Soil properties are often highly correlated among themselves because of the processes of soil development (Moore et al. 1993). As a result, one of the problems encountered when using regression analysis to examine the relationships between yield and a large number of correlated terrain and soil variables is that it may be difficult to determine the relative importance and validity of the variables included in the final model. Multivariate analysis techniques, such as principal component analysis (PCA) and factor analysis (FA) (Hair et al. 1987; Ovalles and Collins 1988; Mallarino et al. 1999; van Es et al. 1999; Kaspar et al. 2004) can be used to avoid the problems of multicollinearity by grouping variables that are strongly correlated and then using these groups as independent variables for regression analysis. Also, multivariate techniques partly circumvent the problems created by correlated variables and could facilitate the interpretation of complex relationships (Mallarino et al. 1999).

However, other techniques have been applied to detect the relationships of soil properties and yield production including principal component analysis-regression (PCA-REG) (Cox et al. 2003; Jiang and Thelen 2004; Shukla et al. 2004), factor analysis-regression (FA-REG) (Mallarino et al. 1999; Kaspar et al. 2004; Ayoubi et al. *2*009), project pursuit regression (PPR) (Drummond et al. 2003), state-space analysis (Wendroth et al. 1997), classification and regression tree (CART) (Stewart et al. 2002; Park et al. 2005), boundary line analysis (Kitchen et al. 1999), and artificial neural network (ANN) analysis (Drummond et al. 2003; Kitchen et al. 2003, Park et al. 2005; Miao et al. 2006).

Several studies have been carried out on the use of ANN in analyzing the relationships between crop yield and soil-landscape factors. For example, Kitchen et al. (2003) reported that ANN outperformed other techniques including multiple linear regression (MLR), stepwise multiple quadratic regression (MQR) in establishing relationships between crop productivity and soil characteristics. Park et al. (2005) compared the efficacy of regression tree, general linear model and ANN techniques for predicting maize yield and found that these methods showed different advantages and disadvantages. However, when used together, they may provide more valuable information on crop response, and more reliable crop growth models may result. Liu et al. (2001) used a feed-forward, back propagation ANN to estimate the non-linear relationships between corn yield and 15 soil, weather and management factors and achieved a RMSE of 20% for 60 verification patterns. Miao et al. (2006) used the ANN method to evaluate the relative importance of selected soil, landscape and hybrid seed factors on corn variability in Illinois (USA), and found that the hybrid seed was more important than soil and landscape factors in the precision crop management. Norouzi et al. (2010) used the ANN analysis to identify the most important topographic and soil attributes in undulating hillslope of western Iran. The ANN wheat yield models for the study area resulted in R^2 and root mean square error (RMSE) of 0.95 and 0.022 Mg ha^{-1} for biomass, 0.93 and 0.021 Mg ha^{-1} for grain yield and 0.89 and 0.063 Mg ha^{-1} for grain protein, respectively. The sediment transport index and total nitrogen were identified as the most important soil and topographic attributes influencing the wheat biomass, grain yield and grain protein. Overall, the results indicated that the ANN models could explain 89–95% of the total variability in wheat biomass, grain yield and grain protein content.

Artificial neural network (ANN) is a mathematical tool, which tries to represent low-level intelligence in natural organisms and it is a flexible structure, capable of making a non-linear mapping between input and output spaces (Gorzalczany 2002). Although barley (*Hordeum vulgare* L.) is one of the most important crops that can be grown in the arid and semiarid regions of Asia (Wahbi and Sinclair 2005; Yusefi et al. 2007), and also in northern Iran, no investigations have been made in determining the most important factors influencing barley production using the non-linear statistical approaches and compare the results with those obtained with linear statistical methods.

In an earlier paper (Ayoubi et al. 2009), we reported on the relationships of barley biomass and grain yield with soil characteristics using factor analysis. In this paper, a comparative evaluation is made of the multivariate regression analysis and ANN for predicting barley production from soil properties at an arid site in northern Iran. Therefore, the objectives of this study were: (i) to compare FA-REG and multilayer perceptron (MLP) neural network methods to predict biomass and grain yield of barley, and (ii) to identify important factors affecting the variability in the barley biomass and grain yield using the soil data by sensitivity analysis within a field in north of Iran.

Materials and methods

Description of the study site and sampling

The study was conducted in a field (36°50′ northern latitude and 54°30′ eastern longitude), located about 60 km north of Gorgan, in Golestan province, Iran (Figure 1). The mean annual temperature at the site is 14.9°C. The mean annual precipitation is 360 mm, which mainly falls from November until March. Soils of the study area are developed on river plain sediments and have less than 2% general slope. Generally, the soil texture is silt loam and silty clay loam in the 0–30 cm soil layer; and the soils of the study area are dominantly classified as fine silty, mixed (calcareous), thermic, Typic Natrargids and fine silty, mixed (calcareous), thermic, Typic Haplosalids according to Soil Taxonomy (Soil Survey Staff 2006). The water table depth with high salinity (EC around 13 dS m^{-1}) varied from 1.5–3 m within the field and its variability followed the micro-topographic changes.

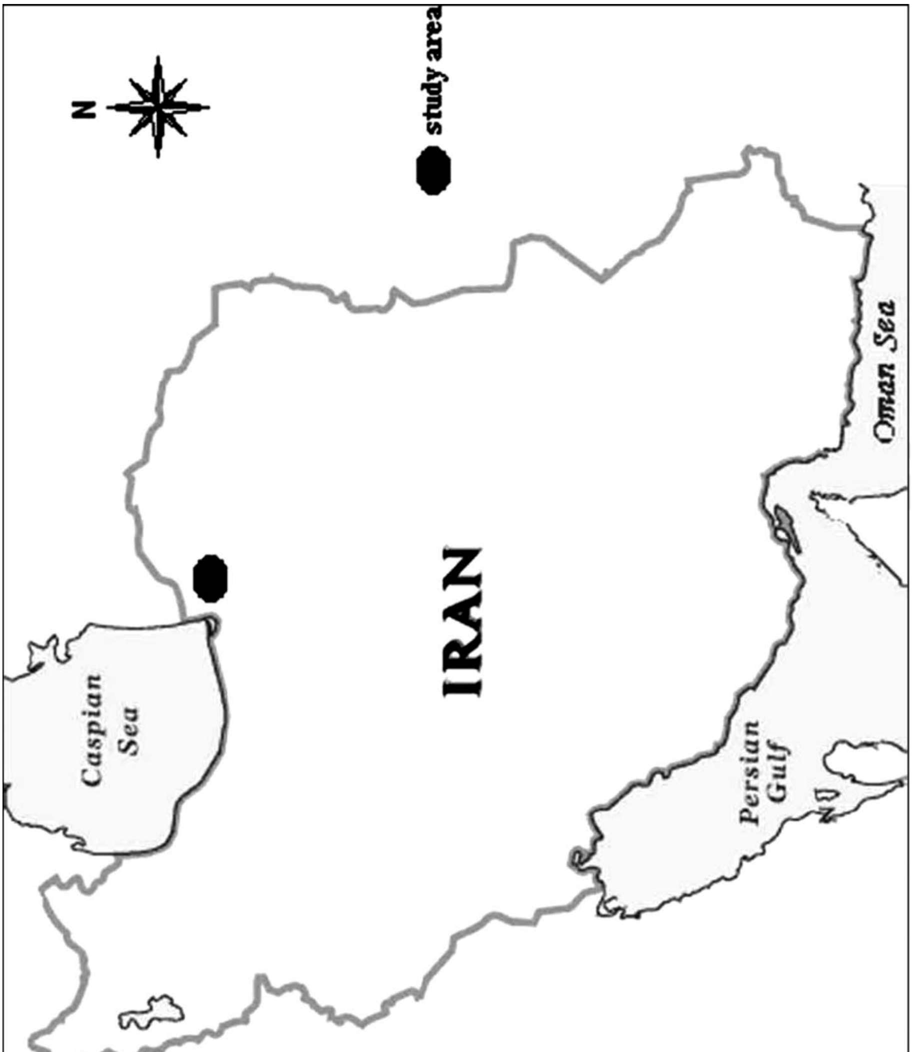

Figure 1. Location of the study area in northern Iran.

As the study area was salt-affected, the field used in the study (5.46 ha in area) had been uncultivated for a long time, and has been partially rehabilitated using surface drainage around the field. Since 2004, the field has been used for barley cultivation. Seedbed preparations included chisel plowing, followed by disking each fall before planting the crop. Fertilizer management consisted of application of 217.39 kg urea ha^{-1}, 131.70 kg KH_2PO_4 ha^{-1}, and 50 kg K_2SO_4 ha in the fall. The barley (cv. Zar) was sown in rate of 300 seeds per m^2, with a 15 cm row spacing with a driller on 15 November 2005.

Soil sampling was performed early July 2006 in 112 selected points in the field according to a grid sampling scheme with 20 $\times$ 30 m distances. The soil sampling at the grid nodes coincided with the harvesting of the crop. Three sub-samples were collected within 1 m^2 area at each grid node, from the top 0–30 cm soil layer, using an auger. Sub-samples were composed to reduce micro-variability. On the same 1 m^2 plots, total aboveground biomass was harvested and grain yield of the barley crop was determined for each sample collected, by separating grains from the chaff. The biomass and grain yield results are expressed on an oven dry basis (for details, see Ayoubi et al. 2009).

Laboratory analysis

The soil samples were air-dried and passed through a 2 mm sieve before using them for chemical and selected physical characteristics. Soil bulk density was measured by core method. The soil samples were oven-dried at 105°C for 24 h and weighted to calculate bulk density (Blake and Hartge 1986). Soil pH was measured in saturated soil using pH electrode (Mclean 1982) and electrical conductivity (EC) was measured in the saturated extract using conductivimeter (Rhoades 1982). Calcium carbonate equivalent (CCE) was measured by Bernard's calcimetric method (Salinity Laboratory Staff 1954). Organic matter (OM) was determined using a wet combustion method (Nelson and Sommers 1982) and total nitrogen (TN) was determined by the Kjeldhal method (Bremner and Mulvaney 1982). Available potassium (K_{ava}) was measured using extraction with ammonium acetate (1N) (Salinity Laboratory Staff 1954); and cation exchange capacity (CEC) was determined by extraction with sodium acetate (Page et al. 1987). Available phosphorous (P_{ava}) was measured by colorimetry using ascorbic acid-ammonium molybdate reagents (Olsen and Sommers 1982). Sodium absorption ratio (SAR) was calculated by measuring Na^+, Mg^{++} and Ca^{++} concentrations in water-extracted solution (Salinity Laboratory Staff 1954).

The wet sieving method of Angers and Mehuys (1993) was used to determine aggregate fractions with a set of sieves of 2.0, 1.0, 0.5, 0.25 and 0.1 mm in diameter. Approximately 50 g of soil sieved through 4.6 mm was put on the first sieve of the set and gently moistened to avoid a sudden rupture of the aggregates. The set was sieved in distilled water at 30 oscillations per minute for 10 min, and the resistant aggregate on each sieve were dried at 105°C for 24 h, weight recorded and corrected for sand fraction, to obtain the proportion of the true aggregates. The mass of <0.1 mm fraction was obtained by difference. The method of van Bavel (1949), as modified by Kemper and Rosenau (1986), was used to determine water stable aggregates (WSA) and mean weight diameter (MWD) of soil aggregates. Available water holding capacity (AWHC) was determined as the difference between field capacity and permanent wilting point (Klute and Dirksen 1986). Water retention at field capacity

(-33 kPa) and at permanent wilting point (-1500 kPa) were determined using high-range pressure plate extractor (Soilmoisture Equipment Corp) equipped with a ceramic plate.

Multivariate regression

Principal component analysis was used to group the 14 soil variables into factors based on the correlation matrix of the variables using PROC FACTOR and the PCA method of factor extraction (Hair et al. 1987; Brejda et al. 2000; SAS Inst. 2000). Principal component analysis was used as the method of factor extraction because it does not require prior estimates of the amount of variation of each soil variable that will be explained by the factors. Its purpose is to derive linear combinations of a set of variables or factors that retain most of the information and variation contained in the variable data set (SAS Inst. 2000).

The maximum number of factors possible is 14, which is equal to the number of variables. Only factors with eigenvalue more than 1 were retained (Hair et al. 1987; Brejda et al. 2000) and were rotated orthogonally with the varimax option (SAS Inst. 2000). Rotation of factors is essentially the application of linear transformation to obtain a more meaningful and discriminating pattern of variable factor loadings within and between factors (Hair et al. 1987). Factor loadings are the correlations of the soil variables with each factor. A stepwise regression procedure (PROC REG; SAS Inst. 2000) was used to regress the barley biomass and grain yield on the factor scores. Selection of the factors for inclusion in the model was based on $p < 0.05$ probability (Freund and Littell 2000; SAS Inst. 2000). Biomass and grain yield were the dependent variables, and the latent variables (Factors) were the independent variables. The modeling was done on training data set including 80% of all samples (90 samples). The selection of the best predictive model was based on the RMSE values and the coefficient of determination (R^2). A data set of 22 samples was used for model validation.

Descriptive statistics in the form of mean, standard error of mean (SE), minimum, maximum, median, coefficient of variation (CV), range, skewness and kurtosis, and also factor analysis and stepwise regression analysis were performed using SPSS (Swan and Sandilands 1995) and SAS (SAS Inst. 2000) software.

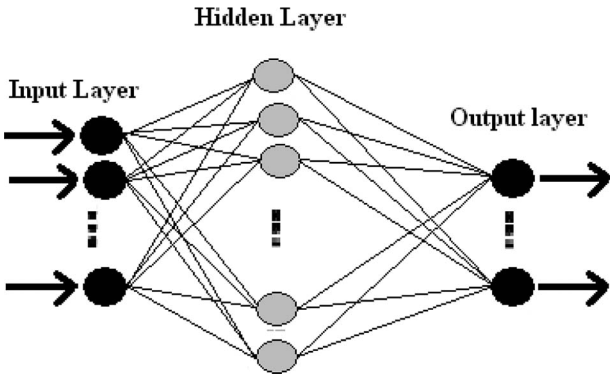

Figure 2. Multilayer artificial neural network (ANN) perceptron neural network learning in ANNs.

Artificial neural network analysis

Neural network processing is based on the performance of many simple processing units called neuron, cell or node. Each neuron in each layer is connected to all elements in the previous and the next layer with links, and each has an associative weight. The general ability of an ANN is to learn and to simulate the natural and complex phenomena (Gorzalczany 2002).

If we consider the $X = [x_1, x_2 \ldots, x_n]$ as input vector and $W = [w_1, w_2 \ldots, w_n]$ as network parameter (weight) vector and if the goal is approximating the multivariate function $f(x)$, the learning procedure is to find the best weight vector (W) to have the best approximation of the $f(x)$. In this paper, multilayer perceptron (MLP) with back propagation learning rule is used.

The MLP network (sometimes called back propagation [BP] network) shown in Figure 2, is probably the most popular ANN used in the engineering problems in the case of non-linear mapping, and it is called 'universal approximator'. The learning process is performed using the BP algorithm (Rumelhart and McClelland 1986). In this study, the standard BP algorithm based on the delta learning rule was used.

Two main processes are performed in a BP algorithm, a forward pass and a backward pass. In the forward pass, an output pattern is presented to the network and its effect propagated through the network, layer by layer. For each neuron, the input value is calculated as follows:

$$net_i^n = \sum_{j=1}^{m} w_{ji}^n \cdot O_j^{n-1} \tag{1}$$

where: net_i^n is the input value of ith neuron in nth layer; w_{ji}^n is the connection weight between ith neuron in nth layer and jth neuron in the $(n\text{-}1)$th layer; O_j^{n-1} is the output of jth neuron in the $(n\text{-}1)$th layer; and m is the number of neurons in the $(n\text{-}1)$th layer.

In each neuron, the value calculated from Equation (1) is transferred by an activation function. The common function for this purpose is the sigmoid function, given by:

$$Sig\left(net_j^n\right) = 1/\left(1 + Exp\left(-net_j^n\right)\right) \tag{2}$$

The output of each neuron is computed and propagated through the next layer until the last layer. Then, the final computed output of the network is compared with the target output. In this regard, an appropriate objective function such as the sum of square error (SSE) or the RMSE is calculated as follows:

$$SSE = \sum_{i=1}^{n_p} \sum_{j=1}^{n_o} \left(T_{pj} - O_{pj}\right)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_p} \sum_{j=1}^{n_o} \left(T_{pj} - O_{pj}\right)^2}{n_p \cdot n_o}} \tag{4}$$

where: T_{pj} is the *j*th element of the target output related to the *p*th pattern; O_{pj} is the computed output of *j*th neuron related to the *p*th pattern; n_p is the number of patterns; and n_o is the number of neurons in the output layer.

After calculating the objective function, the second step of the BP algorithm, i.e. the backward process is started by back propagation of the network error to the previous layers. Using the gradient descent technique, the weights are adjusted to reduce the network error by performing the following equation (Rumelhart and McClelland 1986):

$$\Delta w_{ji}^{n}|_{(m+1)} = \eta \frac{\partial(E)}{\partial w_{ji}^{n}} + \alpha \Delta w_{ji}^{n}|_{(m)} \tag{5}$$

where: $\Delta w_{ji}^{n}|_{(m+1)}$ is the weight increment at the $(m + 1)$th iteration (Epoch); η is the learning rate (Rumelhart and McClelland 1986); α is the momentum term ($0 \leq \eta$, $\alpha \leq 1$) and E is the network error.

This process is continued until the allowable network error is obtained. For designing the artificial neural network, the field data were used. The number of available records collected for this study is 112. The records were shuffled; 72 records were used for the learning process, 20 records were used for validation and 20 sets were used for test, respectively. The network was designed with 14 parameters as input pattern and the biomass or grain yield as the output pattern. Two networks were designed for the estimation of biomass and grain yield separately. The configuration of the designed neural networks for both biomass and grain yield is shown in Figure 3a and 3b, respectively. In this study, ANN models were performed using MATLAB software package (MATLAB version 7.6 with neural network toolbox).

In order to identify the most important soil properties affecting biomass and grain yield of the barley crop, sensitivity analysis was done using the StatSoft method (StatSoft 2004). A relative sensitivity coefficient was calculated as the ratio of the total network error with and without the presence of the given variable. A ratio greater than 1.0 implied that then the variable made an important contribution to the variability in the barley yield components. The higher this ratio the greater is the importance of the variable (StatSoft 2004; Miao et al 2006).

Performance of the methods

A comparison of the two different modeling methods, in order to select the better predictive model, was performed and it was based on the values of the coefficient of determination (R^2) and the root mean square error (RMSE) calculated by similar validation data set (Douaoui et al. 2006).

Results and discussion

Statistical analysis and modeling

The descriptive statistics of the 14 soil chemical and physical parameters and biomass and grain yield of barley crop are presented in Table 1. Skewness values (Table 1) confirmed that all variables were normally distributed. Although a uniform management regime was implemented for this field by the farmer, the variations in both soil parameters and barley biomass and grain yield were considerable (Table 1).

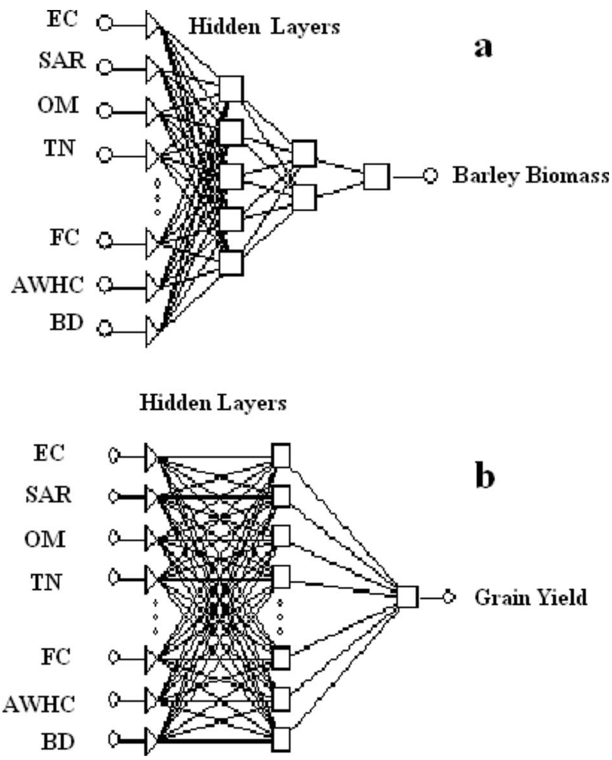

Figure 3. Multilayer perceptron neural network for estimation of biomass (a) and grain yield (b) for the data set in this study. EC, electrical conductivity; SAR, sodium absorption ratio; OM, organic matter; TN, total nitrogen; FC, field capacity; AWHC, available water holding capacity; BD, bulk density.

As a result of the lack of application of plant nutrients through fertilization for a long time, the average available P was 44% lower, and available K was 21% lower compared with the K and P status of fields that have been reclaimed and cultivated for a longer time (>10 years) with wheat and barley (Ayoubi et al. 2007). Electrical conductivity ranged from 1.09–77.3 dS m^{-1} and SAR varied from 1–33, implying that the reclamation of the field was only partially achieved.

In general, the soil properties had CV > 0.35. The most variable in the field were in the order: EC (0.73) $>$ MWD (0.55) $>$ P_{ava} (0.47) $>$ CCE (0.35). The highest CV for EC reflects the presence of micro-topography, which controls the underground water depth and salt accumulation (Kovda 1977). The variables contributing to CV between 0.35 and 0.15, including SAR, CEC, OM and WSA are classified as moderately variable, according to Wilding (1985). The remaining variables indicated low variability (CV < 0.15). Soil pH showed the least CV (0.02) within the field. Several reports confirm the lowest variability in soil pH that occurs within landscapes units of a few ha or less (Cox et al 2003; Shukla et al. 2004).

The results of the FA on soil variables are reported in Table 2. Of the 14 possible factors, only the first five had eigenvalues >1.0 (Table 2). The factors with eigenvalue >1 were retained, since eigenvalue less than one indicated that the factor could explain less variance than could the individual attributes (Sharma 1996; Shukla et al. 2004, 2006).

Table 1. Summary statistics for the selected physical and chemical soil attributes and biomass and grain yield of barley in the field of study ($n = 112$).

Variable	Unit	Mean	SE	Median	CV	Skewness	Kurtosis	Minimum	Maximum	Range
Biomass	$Mg\ ha^{-1}$	2.70	0.09	2.69	0.36	0.39	0.36	0.69	5.23	4.54
Grain yield	$Mg\ ha^{-1}$	0.61	0.02	0.65	0.46	0.04	−0.48	0.10	1.39	1.29
OM	$g\ kg^{-1}$	19.60	0.34	19.4	0.19	0.15	−0.23	10.7	28.00	17.30
EC	$dS\ m^{-1}$	24.39	1.67	19.05	0.73	0.9	−0.26	1.09	77.3	76.2
pH	–	7.68	0.01	7.63	0.02	1.28	3.67	7.36	8.53	1.17
CCE	$g\ kg^{-1}$	212.2	7.05	201.3	0.35	0.41	−0.51	44.9	390	345
MWD	mm	0.37	0.02	0.37	0.55	0.83	1.12	0.01	1.06	1.05
P(ava)	$mg\ kg^{-1}$	15.00	0.67	15.84	0.47	0.13	−0.78	3.00	30.00	27.00
K(ava)	$mg\ kg^{-1}$	346.94	2.22	348.49	0.07	0.31	−0.21	301.00	406.00	105.00
TN	$g\ kg^{-1}$	1.42	0.01	1.40	0.11	1.00	2.44	1.10	2.10	1.00
CEC	$cmol(+)\ kg^{-1}$	17.39	0.35	16.58	0.21	1.41	3.82	11.3	35.00	23.70
SAR	–	21.03	0.45	21.21	0.22	−0.79	2.38	1.00	33.00	32.00
BD	$g\ cm^{-3}$	1.53	0.01	1.54	0.06	0.09	0.02	1.28	1.81	0.53
AWHC	cm	5.39	0.07	5.28	0.14	0.43	−0.59	3.99	7.00	3.00
FC	% Vol	0.19	0.001	0.19	0.09	0.66	−0.19	0.17	0.25	0.08
WSA	%	29.85	0.46	28.77	0.16	0.79	0.93	21.00	45.00	24.00

CV, coefficient of variation (the ratio of the standard deviation to the mean); OM, organic matter; EC, electrical conductivity; CCE, calcium carbonate equivalent; P(ava), available phosphorus; K(ava), available potassium; TN, total nitrogen; CEC, cation exchange capacity; SAR, sodium absorption ratio; BD, bulk density; AWHC, available water holding capacity (for 0–30 cm topsoil); FC, field capacity (for 0–30 cm topsoil); WSA, water stable aggregates. pH, pH in saturated soil with water.

Table 2. Initial eigenvalues, proportional and cumulative variances of first five rotated factors in the selected field in north of Iran.

	Factors				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Initial eigenvalue	3.21	2.21	1.71	1.43	1.09
Proportional variance explained %	30.20	19.71	13.61	7.78	6.50
Cumulative variance explained %	30.20	49.91	63.52	71.30	77.80

The first five factors accounted for 77.8% of total variability of the measured data. The first and the most important factor (F1) explained 30.2% of the total variance. The second factor accounted for a further 19.71% of total variance. The factors 1 and 2 together accounted for 49.91% of the total variance. The inclusion of the three next factors increased the cumulative variance by 27.89% up to 77.8% (Table 2).

The loadings of factors (results are not shown) indicated that factor 1 had the highest loadings for EC and SAR (Ayoubi et al. 2009). Factor 2 showed high positive loadings for P_{ava}, TN and high negative loading for pH. Factor 3 involved the high positive loading for MWD and WSA. Factor 4 associated with FC and AWHC had high positive loadings (Ayoubi et al. 2009). Overall, using the factor analysis, the results showed that the most important soil properties that accounted for most variability were: EC, SAR, pH, TN, P_{ava}, MWD, and WSA.

A stepwise regression analysis was then performed with barley grain yield (Y_g) and biomass yield (Y_b) as dependent variables and scoring coefficient of factors at each sampling location as the independent variables. Selected factors were F1, F2 and F4 for grain yield, and F1, F2 and F3 for biomass, respectively. The regression coefficients were:

$$\mathrm{Yg} = 0.82 - 0.24\,\mathrm{F1} + 0.10\,\mathrm{F2} + 0.07\,\mathrm{F4}\left(R^2 = 0.73; P < 0.001\right) \tag{6}$$

$$\mathrm{Y_b} = 2.53 - 0.82\mathrm{F1} + 0.41\mathrm{F2} + 0.15\,\mathrm{F3}\left(R^2 = 0.79;\ P < 0.001\right) \tag{7}$$

The best ANN models were selected, on the basis of data structure and parameter values, to predict the barley biomass and grain yield. Each of the trained models had 14 input nodes and one output node. The hidden-layer nodes determined were 39 and 41 for biomass and grain yield, respectively. Also, the optimum iteration learning rates determined were 6000 and 7000 for biomass and grain yield, respectively. Normalized predicted data versus normalized observed data for testing data set were plotted (Figure 4a and 4b) for grain yield and biomass, respectively, and the coefficients of determination (R^2) were determined.

The validation analysis showed that the ANN models gave a higher coefficient of determination and lower RMSE than the regression model based on FA. The ANN models resulted in R^2 and RMSE values of 0.93, 0.016 Mg ha^{-1} for biomass, and 0.89, 0.009 Mg ha^{-1} for grain yield, respectively. On the other hand, the FA-REG approach resulted in R^2 and RMSE values of 0.79, 0.047 Mg ha^{-1} for biomass, and 0.73, 0.018 Mg ha^{-1} for grain yield, respectively. The application of ANN produced more powerful models to enable a more precise prediction of barley biomass and

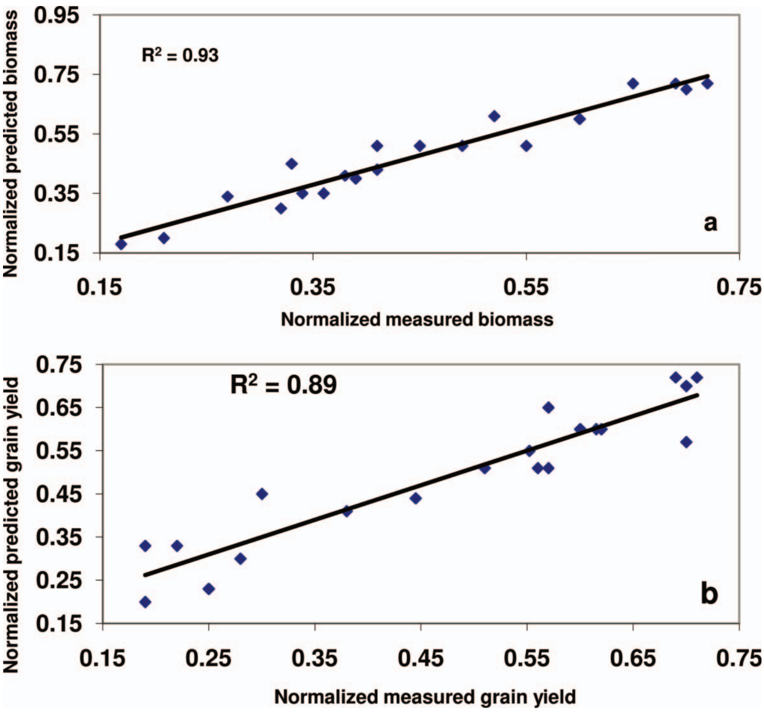

Figure 4. Relationship between values of barley biomass (a) and grain yield (b) as measured and predicted by the ANN models in the study field in northern Iran.

grain yield within the field studied. These results are in agreement with the findings of other researchers (Kitchen et al. 2003; Kaul et al. 2005).

The performance of ANN to a linear statistic method (FA-REG) is a proxy of the non-linear relationships between the soil variables and crop yield. In such cases, the correlation study may provide inaccurate and even misleading results about the relationships (Liu et al. 2001).

Important soil factors influencing barley yield

The relative importance of soil properties using sensitivity analysis based upon coefficients of sensitivity of the ANN models for yield components is shown in Figure 5. The soil variables with higher sensitivity coefficients made an important contribution in explaining the variability of barley yield components.

Electrical conductivity was identified as the most important factor for barley biomass (Figure 5a) and showed the greatest relative coefficient of sensitivity (2.55). Other important factors included SAR, pH, TN, P_{ava}, OM, MWD and WSA with relative coefficients of sensitivity in ranking as 2.25, 2.18, 1.89, 1.84, 1.74, 1.34 and 1.29, respectively. Barley biomass showed lower sensitivity to other soil properties. The relative importance of soil properties, which affected barley biomass, is nearly similar to the results obtained by the FA for the site (Ayoubi et al. 2009).

EC was identified as the most important factor affecting the grain yield (Figure 5b). Other important factors for the grain yield included SAR, pH, OM, TN, P_{ava},

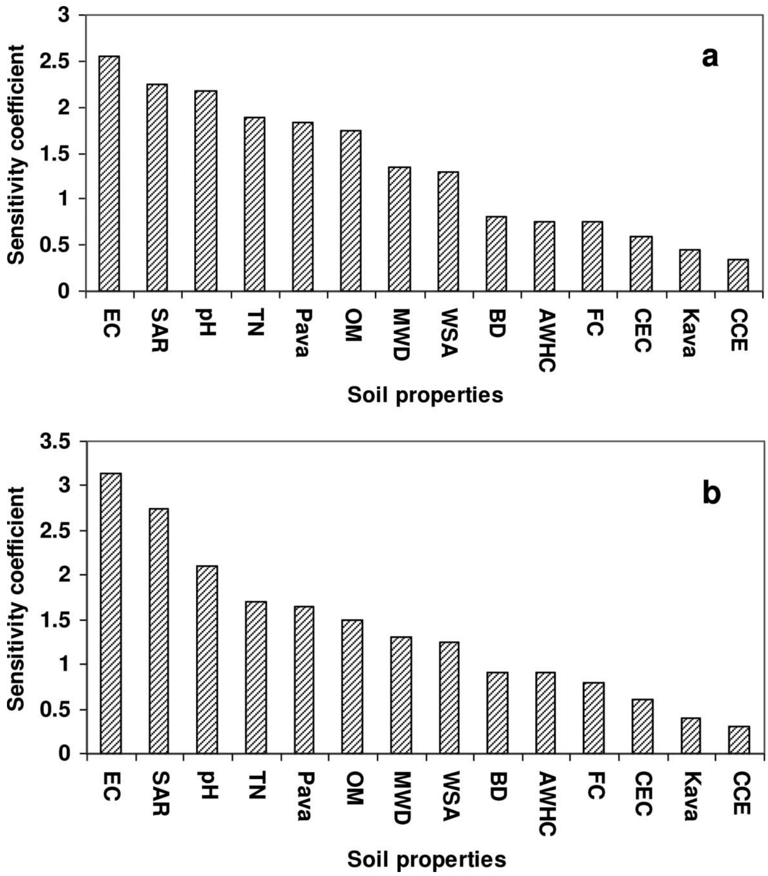

Figure 5. Relative coefficients of sensitivity to soil properties of the ANN models developed for the site in north of Iran. Barley biomass (a) and grain yield (b) as dependent variables. OM, organic matter; EC, electrical conductivity; CCE, calcium carbonate equivalent; P(ava), available phosphorus; K(ava), available potassium; TN, total nitrogen; CEC, cation exchange capacity; SAR, sodium absorption ratio; BD, bulk density; AWHC, available water holding capacity; FC, field capacity; WSA, water stable aggregates; MWD, mean weight diameter.

MWD and WSA. Grain yield showed less sensitivity to the remaining factors and therefore they explained lower variability in grain yield within the studied field. By multivariate analysis, the most important soil properties identified were EC, SAR, pH, TN, P_{ava}, AWHC, and FC (Ayoubi et al. 2009). The relative importance of soil characteristics determined by the ANN analysis is similar to that determined using the factor analysis, with the exception of the relationship of barley biomass and grain yield with soil organic matter. In ANN analysis, OM was included among the most important factors; whereas in the case of multivariate analysis, soil OM was excluded among the important factors to explain grain yield and biomass variability of barley crop. This discrepancy between the results by the two methods probably is a consequence of non-linear relationships between soil organic matter and other soil properties such as P_{ava}, TN, WSA, MWD, FC and AWHC.

EC, SAR, pH, OM, P_{ava}, and TN were consistently identified by ANN as the most important factors for barley biomass and grain yield. The results of the sensitivity analysis for ANN models of barley biomass and grain yield also suggested that the salinity and sodicity had a large effect on the yield variability at the study site. Salinity affects plant growth and production in various ways including: increased water stress, reducing the osmotic potential, adverse effects on the availability of plant nutrients including macro and micronutrients, salt toxicity (Mass and Grieve 1987; Corwin and Lesch 2003; Khan et al. 2004); and sodicity destroys soil physical characteristics (Felhendler et al. 1974; Agassi et al. 1981; Eltaif and Gharaibeh 2007).

In the semi-arid and arid regions, soil water is the major limiting factor for crop production and the processes that control soil water movement and distribution also control crop production (Afyuni et al. 1993; Si and Farrell 2004). Therefore, water availability and its indicators such as FC and AWHC affect barley grain yield.

Organic matter (OM), total nitrogen (TN) and available phosphorus (P_{ava}) were also identified as the three common most important factors affecting barley biomass and grain yield at the site studied. In an agricultural ecosystem, soil organic matter and total nitrogen are the major determinants and indicators of soil fertility and quality, and are closely related to soil productivity (Reeves 1997; Susanne and Michelle 1998; Al-Kaisi et al. 2005; Huang et al. 2007). Furthermore, organic matter has a major role in soil aggregation and soil water retention capacity (Celik 2005). Therefore, it seems that in the field studied, organic matter in parts of the field with low salinity and sodicity improved soil aggregation and water holding capacity of the soil.

Among the selected soil physical characteristics, MWD and WSA of the soil aggregates contributed to the variability in barley yield components. MWD as an indicator of soil aggregates also had significant effects on crop yield variability, especially by controlling the water holding capacity.

The ANN models developed for predicting the barley yield components in the present study explained 89–93% of the total yield variability within the field. Probably, the rest of the variability may be controlled by factors such as soil micronutrient status and factors related to crop, water and soil management especially irrigation management and weed control. Therefore, it can be stated that soil properties combined with management-related factors could further improve the prediction of barley yield variability in the study area. The performance of the ANN models as compared to other approaches appears to have a better ability to predict yield prediction, especially when the complex non-linear relationships exist among factors.

Conclusion

This study was conducted to compare the ability of linear and non-linear functions to predict barley biomass and grain yield from soil properties at the field scale in the arid region of Iran. The results indicated that the ANN models were more powerful tools rather than multivariate regression analysis (FA-REG) to establish the relationships between the soil properties and barley yield components.

Furthermore, the results of this study indicated that production under dry land farming systems in Iran is limited by moisture shortage and the lack of optimum quantity of soil organic matter and the use of improper management practices such as burning of the crop and organic residues. Therefore, the improvement in soil organic matter (SOM) pool, soil aggregation, soil fertility and drainage system could increase barley biomass and grain yield in the study area.

Seven to 11% of the variability of barley yield components remained unexplained using the ANN models. Unexplained variability could have been affected by other factors such as crop, water and soil management practices and the availability of micronutrients, not quantified in the study. Such information on spatial variability in the barley yield is useful for developing site-specific management practices in this region.

Our results also suggest that the ANN analysis can be applied to other crops under similar soil and agro-climatic conditions for predicting crop yield. The use of ANN modeling at additional sites with greater variability in soil properties and soil and crop management practices should help broaden the usefulness of the ANN-based models for the prediction of crop yields.

References

Adams ML, Zhao FJ, McGrath SP, Nicholson FA, Chambers BJ. 2004. Predicting cadmium concentration in wheat and barley grain using soil properties. J Environ Qual. 33:532–541.
Afyuni MM, Cassel DK, Robarge WP. 1993. Effect of landscape position on soil water and corn silage yield. Soil Sci Soc Am J. 57:1573–1580.
Agassi M, Shainberg I, Morin J. 1981. Effect of electrolyte concentration and soil sodicity on infiltration rate and crust formation. Soil Sci Soc Am J. 48:848–851.
Al-Kaisi MM, Yin XH, Licht MA. 2005. Soil carbon and nitrogen changes as influenced by tillage and cropping systems in some Iowa soils. Agric Ecosyst Environ. 105:635–664.
Angers DA, Mehuys GR. 1993. Aggregate stability to water. In: Carter MR, editor. Soil sampling and methods of analysis. Boca Raton (FL): Lewis Publishers. p. 651–657.
Ayoubi S, Khormali F, Sahrawat KL. 2009. Relationships of barley biomass and grain yields to soil properties within a field in the arid region: Use of factor analysis. Acta Agric Scand:. B Soil Plant Sci. 59:107–117.
Ayoubi S, Mohammad Zamani S, Khormali F. 2007. Spatial variability of some soil properties for site-specific farming in northern Iran. Int J Crop Prod. 1(2):225–236.
Blake GR, Hartge KH. 1986. Bulk density. In: Klute A, editor. Methods of soil analysis. Part I, ASA, Monograph, No. 9. Madison (WI): ASA. p. 363–376.
Brejda JJ, Moorman TB, Karlen DL, Dao TH. 2000. Identification of regional soil quality factors and indicators: I. Central and southern high plains. Soil Sci Soc Am J. 64:2115–2124.
Bremner JM, Mulvaney CS. 1982. Nitrogen total. In: Page AL, editor. Methods of soil analysis. Part II. Chemical and microbiological properties. 2nd ed. ASA, Agronomy Monograph, No. 9. Madison (WI): ASA. p. 595–624.
Celik I. 2005. Land use effects on organic matter and physical properties of soil in a southern Mediterranean highland of Turkey. Soil Till Res. 83:270–277.
Corwin DL, Lesch SM. 2003. Application of soil electrical conductivity to precision agriculture. Agron J. 95:455–471.
Cox MS, Gerard PD, Wardlaw MC, Abshire MJ. 2003. Variability of selected soil properties and their relationships with soybean yield. Soil Sci Soc Am J. 67:1296–1302.
Douaoui AEK, Nicolas H, Walter C. 2006. Detecting salinity hazards within a semi-arid context by means of combining soil and remote sensing data. Geoderma. 134:217–230.
Drummond ST, Sudduth KA, Joshi A, Birrell SJ, Kitchen NR. 2003. Statistical and neural methods for site-specific yield prediction. Trans ASAE. 46(1):5–14.
Eltaif NI, Gharaibeh MA. 2007. Effects of single and mixed ion solutions on hydraulic and physical properties of a clay soil. Water, Air, Soil Pollut. 181:297–302.
Felhendler R, Shainberg I, Frenkel H. 1974. Dispersion and hydraulic conductivity of soils in mixed solution. Trans 10th Int Congress Soil Sci Moscow. 1:103–112.
Freund RJ, Littell RC. 2000. SAS system for regression. Cary (NC): SAS Inst.
Godwin RJ, Miller PCH. 2003. A review of the technologies for mapping within-field variability. Biosyst Eng. 84:93–407.
Gorzalczany MB. 2002. Computational intelligence systems and applications. Heidelberg, Germany: Physica-Verlag Company.

Hair JF, Anderson RE, Tatham RL. 1987. Multivariate data analysis with readings. New York: Macmillan Publishing Company. p. 449.

Huang B, Sun WX, Zhao YC, Zhu J, Yang RQ, Zou Z, Ding F, Su JP. 2007. Temporal and spatial variability of soil organic matter and total nitrogen in an agriculture ecosystem as affected by farming practices. Geoderma. 139:336–345.

Jiang P, Thelen KD. 2004. Effects of soil and topographic properties on corn yield in a north-central corn-soybean cropping system. Agron J. 96:252–258.

Kaspar TC, Pulido DJ, Fenton TE, Colvin TS, Karlen DL, Jaynes DB, Meek DW. 2004. Relationship of corn and soybean yield to soil and terrain properties. Agron J. 96:700–709.

Kaul M, Hill R, Walthall C. 2005. Artificial neural networks for corn and soybean yield prediction. Agric Syst. 85:1–18.

Kemper WD, Rosenau RC. 1986. Aggregate stability and size distribution. In: Klute A, editor. Methods of soil analysis. Part I. ASA, 2nd ed. Monograph, No. 9. Madison (WI): ASA. p. 687–734.

Khakural BR, Robert PC, Huggins DR. 1999. Variability of corn/soybean yield and soil landscape properties across a southwestern Minnesota landscape. In: Robert PC, et al. Proc 4th Int Conf Precision Agriculture, St Paul, MN, USA; 1998, July 19–22; Madison (WI): ASA, CSSA, and SSSA. p. 573–579.

Khan A, Khan A, Nisar Ahmad MZ. 2004. Effect of salinity and sodicity on crop production at Kot Kashmir (Lakki Marwat). Sar J Agric. 20(4):583–590.

Kitchen NR, Sudduth KA, Drummond ST. 1999. Electrical conductivity as a crop productivity measure for claypan soils. J Product Agric. 12(4):607–617.

Kitchen NR, Drummond ST, Lund ED, Sudduth KA, Buchleiter GW. 2003. Electrical conductivity and topography related to yield for three contrasting soil-crop systems. Agron J. 95:483–495.

Klute A, Dirksen C. 1986. Hydraulic conductivity and diffusivity. In: Klute A, editor. Methods of soil analysis. Part 1. 2nd ed. Agronomy monograph, Vol. 9. Madison (WI): ASA. p. 687–734.

Kovda VA. 1977. Arid land irrigation and soil fertility: Problems of salinity, alkalinity, and compaction. In: Worthington EB, editor. Arid land irrigation in developing countries: Environmental problems and effects. Oxford: Pergamon Press. p. 211–236.

Kravchenko AN, Bullock DG. 2000. Correlation of corn soybean grain yield with topography and soil properties. Agron J. 92:75–83.

Liu J, Goering CE, Tian L. 2001. A neural network for setting target yields. Transact ASAE. 44(3):705–713.

Mallarino AP, Oyarzabal ES, Hinz PN. 1999. Interpreting within-field relationships between crop yields and soil and plant variables using factor analysis. Precision Agric. 1:15–25.

Mass EV, Grieve CM. 1987. Crop salt tolerance current assessment. J Irrig Drainage Div ASCE. 103:559–564.

Mclean EO. 1982. Soil pH and lime requirement. In: Page AL, editor. Methods of soil analysis. Part II, 2nd ed. ASA, Monograph. Madison (WI): ASA. p. 199–223.

Miao Y, Mulla DJ, Robert PC. 2006. Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. Precision Agric. 7:117–135.

Moore ID, Gessler PE, Nielson GA. 1993. Soils attribute prediction using terrain analysis. Soil Sci Soc Am J. 57:443–452.

Nelson DW, Sommers LE. 1982. Total carbon, organic carbon, and organic matter. In: Buxton DR, editor. Methods of soil analysis. Part II, 2nd ed. ASA, Monograph, No. 9. Madison (WI): ASA. p. 539–579.

Norouzi M, Ayoubi A, Jalalian A, Khademi H, Dehghani AA. 2010. Predicting rainfed wheat quality and quantity by artificial neural network using terrain and soil characteristics. Acta Agric Scand B: Soil Plant Sci. 60:341–352.

Olsen SR, Sommers LE. 1982. Phosphorus. In: Page AL, editor. Methods of soil analysis. Agronomy No. 9, Part 2: Chemical and microbiological properties. 2nd ed. Madison (WI): ASA. p. 403–430.

Ovalles FA, Collins ME. 1988. Variability of northwest Florida soils by principle component analysis. Soil Sci Soc Am J. 52:1430–1435.

Page MC, Sparks DL, Noll M, Hendricks GJ. 1987. Kinetics and mechanisms of potassium release from sandy middle Atlantic Coastal plain soils. Soil Sci Soc Am J. 51:1460–1465.

Park SJ, Hwang CS, Vlek PLG. 2005. Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions Agric Syst. 85:59–81.
Quine TA, Zhang Y. 2002. An investigation of spatial variation in soil erosion, soil properties and crop production within an agricultural field in Devon, UK. J Soil Water Cons. 57: 50–60.
Reeves DW. 1997. The role of soil organic matter in maintaining soil quality in continuous cropping systems. Soil Till Res. 43:131–167.
Rhoades JD. 1982. Soluble salts. In: Page AL, editor. Methods of soil analysis. Part II. 2nd ed. ASA, Monograph, No. 9. Madison (WI): ASA. p. 167–179.
Rumelhart DE, McClelland JL, PDP Research Group. 1986. Parallel recognition in modern computers. In: Proceedings: Explorations in the microstructure of Cognition. Cambridge, MA: MIT Press/Bradford Books. Foundations, Vol 1.
Salinity Laboratory Staff. 1954. Diagnosis and improvement of saline and alkali soils. No. 60. Washington (DC): USDA.
SAS Inst. 2000. SAS 9.1.3. Help and Documentation. Cary (NC): SAS Institute Inc.
Sharma S. 1996. Applied multivariate techniques. New York: John Wiley and Sons.
Shukla MK, Lal R, Ebinger M. 2004. Principal component analysis for predicting corn biomass and grian yields. Soil Sci. 169:215–224.
Shukla MK, Lal R, Ebinger M. 2006. Determining soil quality indicators by factor analysis. Soil Till Res. 87:194–204.
Si BC, Farrell RE. 2004. Scale dependent relationship between wheat yield and topographic indices: A Wavelet approach. Soil Sci Soc Am J. 68:577–587.
Soil Survey Staff. 2006. Keys to soil taxonomy. U.S. Department of Agriculture, Natural Resources Conservation Service.
StatSoft Inc. 2004. Electronic statistics textbook (Tulsa, OK, USA). http://www.statsoft.com/textbook/stathome.html
Stewart CM, McBratney AB, Skerritt JH. 2002. Site-specific durum wheat quality and its relationship to soil properties in a single field in Northern New South Wales. Precision Agric. 3:155–168.
Susanne A, Michelle MW. 1998. Long-term trends of corn yield and soil organic matter in different crop sequences and soil fertility treatments on the Morrow Plot. Adv Agron. 62:153–197.
Swan ARH, Sandilands M. 1995. Introduction to geological data analysis. London: Blackwell. p. 446.
van Bavel CHM. 1949. Mean weight diameter of soil aggregate as a statistical index of aggregation. Soil Sci Soc Am Proc. 14:20–23.
van Es HM, Ogden CB, Hill RL, Schindelbeck RR, Tsegaye T. 1999. Integrated assessment of space, time, and management-related variability of soil hydraulic properties. Soil Sci Soc Am J. 63:1599–1608.
Vieira SR, Paz Gonzalez A. 2003. Analysis of the spatial variability of crop yield and soil properties in small agricultural plots. Bragantia, Campinas. 62:127–138.
Wahbi A, Sinclair TR. 2005. Simulation analysis of relative yield advantage of barley and wheat in an eastern Mediterranean climate. Field Crop Res. 91:287–296.
Wendroth O, Reynold WD, Vieira SR, Reichardt K, Wirth S. 1997. Statistical approaches to the analysis of soil quality data. In: Gregorich EG, Carter MR, editors. Soil quality for crop production and ecosystem health. Amsterdam: Elsevier. p. 247–276.
Wilding LP. 1985. Spatial variability: Its documentation, accommodation, and implication to soil surveys. In: Nielsen DR, Bouma J, editors. Soil spatial variability. Wageningen, The Netherlands: Pudoc.
Yusefi S, Wissal M, Mahmoudi H, Abdelly C, Gharsalli M. 2007. Effect of salt on physiological responses of barley to iron deficiency. Plant Phys Biochem. 45:309–314.