Acta Agriculturae Scandinavica Section B – Soil and Plant Science, 2009; 59: 107–117

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Relationships of barley biomass and grain yields to soil properties within a field in the arid region: Use of factor analysis

SHAMSOLLAH AYOUBI[1], FARHAD KHORMALI[2] & KANWAR L. SAHRAWAT[3]

[1]*Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan, Iran,* [2]*Department of Soil Science, College of Agriculture, Gorgan University of Agricultural Sciences and Natural Resources, Gorgan, Iran, and* [3]*International Crop Research Institute for The Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India*

Abstract
Understanding the variability of soil properties and their effects on crop yield is a critical component of site-specific management systems. The objective of this study was to employ factor and multiple regression analyses to determine major soil physical and chemical properties that influence barely biomass and grain yield within a field in the arid region of northern Iran. For this purpose, soil samples and crop-yield data were collected from 108 sites, at regular intervals (20 × 30 m) in a 5.6 ha field. Soil samples were analysed for total nitrogen (TN), available phosphorus (P_{ava}), available potassium (K_{ava}), cation-exchange capacity(CEC), electrical conductivity (EC), pH, mean weight diameter of aggregates (MWD), water-stable aggregates (WSA), field capacity volumetric (FC), available water-holding capacity (AWHC), bulk density (BD), and calcium carbonate equivalent (CCE). Results of the factor analysis, followed by regression of biomass and grain yield of barley with soil properties, showed that the regression equations developed accounted for 78 and 73% of the total variance in biomass and grain yield, respectively. Study of covariance analysis among soil variables using factor analysis indicated that some of the variation measured could be grouped to indicate a number of underlying common factors influencing barley biomass and grain yields. These common factors were salinity and sodicity, soil fertility, and water availability. The most effective soil variables to barley production in the study area identified as EC, SAR, pH, TN, P_{ava}, AWHC, and FC. In this study, factor analysis was effective to identify the groups of correlated soil variables that were significantly correlated with the within field variability in the yield of the barley crop. Our results also suggest that the approach can be applied to other crops under similar soil and agroclimatic conditions.

Keywords: *Barley yield prediction, factor analysis, site-specific management, soil and crop variability.*

Introduction

There is increased interest in the use of precision agriculture to better understand the within-field variability, for efficient management of various fertility-enhancing inputs mostly in high-input modern agriculture but also in shifting cultivation (Godwin & Miller, 2003). Site-specific management is the process of managing soils based on localized conditions within field boundaries (Carr et al., 1991).

Soil properties vary in time and space, and variability of soil properties is the rule rather than the exception. Natural variability of soil results from complex interactions between geology, topography, and climate as well as cultivation, land use, and soil erosion (Quine & Zhang, 2002). The within-field variability in soil properties influences soil processes such as water and nutrient movement and their redistribution and supply to plants, and root growth and sustenance; and the variability also influences crop response to management and the susceptibility of soil to degradation (Shukla et al., 2004). Knowledge of the variability of soil properties is essential to selecting as well as effectively applying management decisions in the field (Vieira & Paz Gonzalez, 2003; Shukla et al., 2004). Determining which soil factors to base management decisions on is often a complex processes due to interactions among various factors that affect crop yield.

Several techniques have been used to understand the relationships between crop yield and soil or

Correspondence: S. Ayoubi, Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan, 84156-83111, Iran. Tel: +98-311-3913343. Fax: +98-311-3913471. E-mail: ayoubi@cc.iut.ac.ir

(Received 15 October 2007; accepted 10 December 2007)

ISSN 0906-4710 print/ISSN 1651-1913 online
DOI: 10.1080/09064710801932417

landscape properties in an attempt to identify important factors that influence the relationships. Correlation and multiple linear regressions (MLR) are commonly used for such purposes (Khakural et al., 1999; Kravchenko & Bullock, 2000; Adams et al., 2004), but the results are often not satisfactory. Soil properties are often highly correlated with each other 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 the difficulty of determining the relative importance and validity of the variables included in the final model. Additionally, variables may be selected for the model even when there is no obvious mechanistic basis for their inclusion because they are strongly correlated with a variable that does have a mechanistic relationship with yield.

The problem caused by correlated variables can be minimized by grouping variables, so that the correlation of two variables from the same group is large. Each group can be represented by a new variable that is created from the variables in the group (Jolliffe, 1986). Multivariate analysis techniques, such as principal component analysis (PCA) and factor analysis (FA) (Hair et al., 1987; Ovalles & 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).

Multivariate analysis in combination with multiple regressions can evaluate combined effects of variation of soil properties on biomass and yield production (Vieira & Paz Gonzalez, 2003; Jiang & Thelen, 2004; Kaspar et al., 2004; Shukla et al., 2004). Factor analysis often is more successful at identifying groups of correlated variables because it is an analysis of covariances whereas principal-component analysis is an analysis of variances. Mallarino et al. (1999) using factor analysis showed that some of the variables measured could be grouped to indicate a number of underlying common factors influencing corn yields in five different fields. The extracted factors included soil fertility, weed control, and conditions for early plant growth. Their results indicated that the choice of site variables to be measured is very important because the variables that may explain yield variability most probably are different across fields. Kaspar et al. (2004), by measurement of 20 soil and terrain variables, tried to determine the relative importance of soil and terrain parameters in soybean- and corn-yield variability. They also compared capability of a 20-variable data set with a seven easily measured terrain properties in order to detect variability of the crops. Factor analysis of data set led to factors termed as 'landscape position', 'closed depression', 'pH', and 'curvature'. Factor analysis of the variables, followed by a regression of yield on the resulting factors, showed that 20-variable data set explained more spatial variation in yield than did the subset of seven variables.

Barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*) are similar crops that can be grown in the arid regions of Asia. However, barley is generally favored over wheat in drier areas (Wahbi & Sinclair, 2005). Barley is one of the major crops in arid and semiarid regions, being relatively resistant to aridity and salinity (Yusefi et al., 2007). Since a great part of Iran (approximately 90% of the country) is arid and semiarid, barley is a common crop in cropping systems. There is little information on factors influencing barley production in arid regions of Iran. So for improvement in management practices, it is necessary to identify which soil properties control barley biomass and grain-yield variability in a field. Although several attempts have been made to predict crop biomass and grain yield by some selected soil physical and chemical properties using PCA or FA and multiple regression, little information is available on prediction of barley biomass and grain yields by factor and regression analyses, especially in arid zones of Iran. Therefore, this study was conducted to determine the variability of selected soil physical and chemical properties at the field scale and to group measured soil properties into a few latent variables (Factors) to explain the variability in barley grain and biomass yields.

Material and methods

Description of the study site and sampling

The study was conducted on a farmer-operated barely field north of Aq Qala city, located about 60 km north of Gorgan in Golestan province, Iran (Figure 1). Since the study area has been salt-affected, the field used in the study (5.46 ha in area) has been uncultivated for a long time (20 years), and currently has been partially rehabilitated using surface drainage. Since 2004, the field has been used for barley cultivation. The mean annual temperature at the site is 14.9°C. The mean annual precipitation is 360 mm which falls mainly from November through March. Soils of the study area are developed on river-plain sediments and have less

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

than 2% general slope. Generally, the soil texture is silty 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) varies from 1.5 to 3 m within the field and its variability follows the microtopography changes.

Seedbed preparations included chisel plowing, followed by disking each fall (autumn) before planting the crop. Considered typical, fertilizer management consisted of application of 100-30-50 kg (N-P-K) in the fall (autumn). Planting of barley (cv. Zar) was at a rate of 300 seeds per m^2, with an 18 cm row spacing with a driller on 15 November, 2005.

Soil sampling were performed early July 2006 in 108 selected points in the field on a grid sampling scheme with 20×30 distances, and the soil sampling coincided with the harvesting of the crop. Soil samples were collected from 0–30 cm depth using an auger, three sub-samples per 1 m^2 area in each site, and then composed to reduce microvariability. 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, and the biomass and grain-yield results are expressed (Mg/ha) on an oven-dry basis.

Soil analysis

Prior to analyses, the soil samples were air-dried under shade for two weeks, after which they were ground to pass through a 2-mm sieve to remove stones, roots, and large organic residues for chemical and selected physical characteristics. Soil bulk density was measured by the core method. The soil samples were oven-dried at 105°C for 24 h and weighed to calculate bulk density (Blake & Hartge, 1986). pH was measured in saturated soil using pH electrode (Mclean, 1982) and electrical conductivity (EC) was measured in the saturated extract using a conductimeter (Rhoades, 1982). Calcium carbonate equivalent (CCE) was measured by Bernard's calcimetric method (Salinity Laboratory Staff, 1954). Soil organic matter (SOM) was determined using a wet-combustion method (Nelson & Sommers, 1982) and total nitrogen (TN) was determined by the Kjeldhal method (Bremner & 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 phosphorus (P_{ava}) was measured by colorimetry using ascorbic acid–ammonium molybdate reagents (Olsen & 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 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 aggregates. The set was sieved in distilled water at 30 oscillations per minute for 10 minutes, and the resistant aggregate on each sieve were dried at 105°C for 24 h. The weight was 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 Bevel (1949), as modified by Kemper and Rosenau (1986), was used to determine water-stable aggregates (WSA) and mean weight diameter (MWD).

The WSA was calculated using Equation (1), where $M_{(a+s)}$ is the mass of resistant aggregates plus sand (g), M_s the mass of the sand fraction alone (g), and M_t the total mass of the sieved soil (g).

$$WSA = \frac{(M_{(a+s)} - M_s)}{(M_t - M_s)} \times 100 \qquad (1)$$

Available water-holding capacity (AWHC) was determined as the difference between field capacity and permanent wilting point (Klute & Dirksen, 1986). Water retention at field capacity (−33 kPa) and at permanent wilting point (−1500 kPa) were determined using a pressure-plate apparatus.

Statistical analysis

Descriptive statistics in the form of mean, standard error of mean (SE), minimum, maximum, median, coefficient of variation (CV), distribution of normality, range, skewness, and kurtosis were determined (Wendroth et al., 1997). The CV was used to describe the amount of variability for each soil parameter and yield. Pearson linear correlations among parameters were calculated using SPSS software (Swan & Sandilands, 1995) and were used to interpret relationships among soil and yield variables.

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 requires no 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 >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 between the soil variables and each factor. A stepwise regression procedure (PROC REG; SAS Inst., 2000) was used to regress barley biomass and grain yield on the factor scores. Selection of factors for inclusion in the model was based on probability ≤0.05 (Freund & Littell, 2000; SAS Inst., 2000).

Biomass and grain yield were the dependent variables, and the latent variables (Factors) were the independent variables. The modelling was done on training data set including 80% of all samples (90 samples).The models were of the form shown in Equation (2), where Y represents estimated barley biomass and grain yields, b_o to b_n are coefficients, F_1 to F_n are the latent variables, and ε represents residual error. The selection of the best predictive model was performed based on root-mean-square error (RMSE) values and determination coefficient (R^2), using validation data set (22 samples). RMSE was determined by Equation (3)(Douaoui et al., 2006), where n donates the number of samples, and Z^* and Z are the predicted and measured values, respectively.

$$Y = b_0 + b_1F_1 + b_2F_2 + \ldots + b_nF_n + \varepsilon \qquad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(Z^* - Z)^2} \qquad (3)$$

Once the factors that were significantly related to biomass and grain yield of barley were found, their meaning in terms of soil components had to be determined. Soil parameters with larger loadings in eigenvectors contribute more to the factor and can be interpreted as the effective properties in variability of yield production.

Results and discussion

Parameter statistics

The descriptive statistics of 14 soil chemical and physical parameters and biomass and grain yield of barley are presented in Table I. All variables were

Table I. Summary statistics for the selected physical and chemical soil attributes and biomass and grain yield of barley in the field of study ($n = 108$).

Variable	Unit	Mean	SE	Median	CV	Skewness	Kurtosis	Minimum	Maximum	Range
Biomass	Mg/ha	2.68	0.09	2.69	0.36	0.33	0.31	0.69	5.23	4.54
Grain yield	Mg/ha	0.61	0.02	0.66	0.46	0.02	0.53	0.10	1.39	1.29
OM	gr/kg	19.63	0.35	19.50	0.19	0.14	−0.24	10.70	28.00	17.30
EC	dS/m	24.25	1.67	19.33	0.70	0.71	−0.63	1.10	63.30	62.2
pH	$-\text{Log}[H^+]$	7.67	0.01	7.64	0.02	0.60	0.35	7.36	8.19	0.83
CCE	gr/kg	211.98	7.05	200.10	0.35	0.44	−0.47	45.00	390	345
MWD	mm	0.38	0.02	0.38	0.53	0.68	0.80	0.01	0.96	0.95
P_{ava}	mg/kg	15.16	0.69	15.90	0.47	0.09	−0.80	3.00	30.00	27.00
K_{ava}	mg/kg	346.85	2.18	349.18	0.07	0.28	−0.25	305.00	401.00	96.00
TN	gr/kg	1.42	0.01	1.40	0.11	0.71	0.81	1.10	1.9	0.80
CEC	Cmol(+)/kg	17.21	0.32	16.54	0.19	0.79	0.49	11.3	26.70	15.40
SAR	–	20.93	0.46	21.19	0.22	−0.79	2.33	1.00	33.00	32.00
BD	gr/cm^3	1.53	0.01	1.54	0.06	0.18	−0.08	1.32	1.81	0.49
AWHC	cm	5.40	0.07	5.29	0.14	0.42	−0.62	4.00	7.00	3.00
FC	% (vol)	0.19	0.001	0.19	0.09	0.75	−0.09	0.17	0.25	0.08
WSA	%	29.71	0.44	29.45	0.16	0.66	0.91	21.00	45.00	24.00

OM: Organic matter; EC: Electrical conductivity; CCE: Calcium carbonate equivalent; MWD: Mean weight diameter; 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.

normally distributed (according to the Kolmogorov–Smirnov test). Skewness values (Table I) also confirmed that all variables were normally distributed, although some researchers suggested that, in disturbed ecosystems, some soil variables show skewed distribution (Wang et al., 2003), but normality test and skewness values of soil properties showed low deviation from normal distribution. Based on the earlier research of the site (field) studied (Khormali & Ayoubi, 2006), the surface-soil texture showed the minimum variability and was classified mainly as Silt or Silty Clay Loam. Although a uniform management regime was implemented for this field by the farmer, variation in both soil parameters and barley production was considerable (Table I). As a result of the lack of fertilization for a long time, average available P was 43.39% 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 1.1 to 63.3 dS/m and SAR varied from 1.00 to 33, implying that reclamation of the field had been only partially achieved; and this was reflected in microvariability in elevation throughout the field; this in conjunction with high evapo-transpiration potential in the study area led to high saline and sodic zones within the field. Owing to high Na content, low levels of OM, and consequently poor soil aggregation, MWD and WSA were 68% and 44% lower than that of cultivated fields with similar soil properties (soil texture, CEC, and clay mineralogy) in the same province (Ayoubi, 2005).

Parameter variability

Soil variability is a key element in site-specific soil management. Variability in space and time for point data can give valuable insight into the dynamic nature of soil properties within a field's boundary. Management of this variability is worthwhile if the variability is high enough to justify the costs for obtaining the information or if the proposed management will increase profit. The knowledge of the variability in soil physical and chemical properties is a key for designing site-specific management practices (Shukla et al., 2004). To define variability into three classes, we applied the system suggested by Wilding (1985). In general the properties having CV > 0.35 and highly variable in the field were in the order of EC (0.70) > MWD (0.53) > P_{ava} (0.47) > CCE (0.35). The highest CV for EC reflects the role of microtopography in controlling the underground water depth and salt accumulation (Kovda, 1977). The variables contributing CVs between 0.35 and 0.15 are classified as moderate variables including SAR, CEC, OM, and WSA. The remaining variables indicated low variability (CV < 0.15). pH showed the least CV (0.02) within the field. Several researchers confirmed the lowest variability for pH that occurs within landscape units of a few hectares or less (Cox et al., 2003; Shukla et al., 2004).

Correlation analysis

The linear correlation analysis of the 14 soil attributes, which represent soil physical and chemical properties for the 0–30 cm depth and barley biomass

Table II. Pearson correlation coefficients for soil physical and chemical properties and barley biomass and grain yield for the field of study ($n = 108$).

	Biomass	Grain yield	OM	EC	pH	CCE	CEC	SAR	P_{ava}	TN	K_{ava}	MWD	BD	AWHC	FC	WSA
Biomass	1															
Grain yield	0.78**	1														
OM	0.43**	0.36**	1													
EC	−0.76**	−0.67**	−0.41**	1												
pH	−0.07	−0.18*	0.07	0.11	1											
CCE	−0.13	−0.12	−0.23*	0.14	−0.31**	1										
CEC	0.13	−0.01	0.06	−0.13	−0.05	−0.42**	1									
SAR	−0.65**	−0.51**	−0.33**	0.55**	0.04	0.01	0.04	1								
P_{ava}	0.23*	0.13	0.01	−0.23*	−0.78**	−0.12	0.01	−0.27**	1							
TN	0.39**	0.13	0.88**	−0.25**	0.66**	0.05	0.32**	−0.26**	0.49**	1						
K_{ava}	0.19*	0.15	−0.12	−0.10	0.37**	−0.12	−0.01	0.19*	0.08	−0.02	1					
MWD	0.33**	0.22*	0.61**	−0.29**	−0.23*	−0.14	−0.02	−0.28**	−0.02	0.20*	−0.01	1				
BD	−0.15	0.04	−0.16	0.06	−0.24*	0.08	−0.32**	0.03	−0.07	−0.19*	−0.03	−0.16	1			
AWHC	0.34**	0.41**	0.45**	−0.40**	−0.25*	0.05	−0.12	−0.29**	0.30**	0.20*	−0.28**	0.42**	−0.01	1		
FC	0.21*	0.19*	0.14	−0.34**	−0.01	0.15	0.22*	−0.33**	0.31**	0.07	−0.48**	0.03	0.02	0.79**	1	
WSA	0.36**	0.22*	0.37**	−0.31**	−0.08	−0.21**	−0.05	−0.20*	0.01	0.15	0.07	0.87**	−0.47**	0.11	0.01	1

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

and grain yield, showed there was significant correlation among 64 of the 120 soil attribute pairs ($P < 0.01$, and $P < 0.05$) (Table II). Barley biomass and grain yield were positively correlated ($r = 0.78$, $P < 0.01$) and barley biomass was also positively correlated with OM, TN, and WSA ($r > 0.35$), AWHC, and MWD ($r > 0.33$) and had low positive correlation with FC, K_{ava}, and P_{ava} ($r > 0.18$, $P < 0.05$). Barley biomass showed negative correlation with EC ($r = -0.76$, $P < 0.001$) and SAR ($r = -0.65$, $P < 0.01$). The grain yield of barley within the field was positively correlated with OM and AWHC ($r > 0.36$, $P < 0.01$), MWD, FC, and WSA ($r > 0.19$, $P < 0.05$), and negatively correlated with EC ($r = -0.67$, $P < 0.001$), SAR ($r = -0.51$), and pH ($r = -0.18$). Shukla et al. (2004) reported positive correlation between corn grain yield and OM, WSA, TN, and AWC (available water capacity) and negative correlation with soil pH. The negative correlation of barley biomass and grain yield with EC, SAR, and pH showed that the salinity and sodicity significantly affected crop production in the study area. Also, positive correlation of AWHC, MWD, FC, and WSA with barley yield indicated that the soil aggregation, which impacts both transmission and storage pores, affected both biomass and grain yield (Shukla et al., 2004).

High positive correlation ($r = 0.88$) of total nitrogen with SOM indicated that mineral nitrogen (NH_4^+ and NO_3^-) used by farmers did not influence variability of total nitrogen, which was controlled mainly by organic nitrogen. Positive correlation between EC and SAR ($r = 0.55$) revealed that a considerable portion of soil solution contained Na^+. MWD and WSA were highly positively correlated ($r = 0.87$, $P < 0.0001$) and their positive correlation with OM (0.61 and 0.37, respectively) and negative correlation with SAR (−0.28 and −0.20, respectively) are indicative of the instructive impact of OM on soil aggregation and the influence of sodicity on soil dispersion. Aggregate stability depends on interaction between primary particles and organic constituents to form stable aggregates, which are influenced by various factors related to soil environmental conditions and management practices (Elustondo et al., 1990; Celik, 2005). Soil organic matter plays a key role in the formation and stabilization of soil aggregate (Lu et al., 1998).

Factor analysis

Factor analysis is the name of a class of multivariate statistical methods that can be used to summarize and describe large groups of variables (Hair et al., 1987; Brejda et al., 2000; Kaspar et al., 2004). It can be used to identify relationships among groups of

Table III. Rotated factor loadings and communalities of measured variables for the first five factors with eigenvalue > 1.

Variable	Factors					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	CE*
OM	−0.21	0.14	0.40	0.12	0.45	0.46
EC	0.87	0.05	−0.17	−0.18	−0.18	0.89
PH	0.14	−0.85	0.01	−0.11	0.09	0.77
CCE	0.25	−0.57	−0.28	0.20	0.28	0.62
MWD	−0.08	−0.01	0.94	0.05	0.08	0.91
CEC	−0.12	−0.09	−0.05	0.04	−0.05	0.83
SAR	0.61	−0.08	−0.34	0.30	0.36	0.75
BD	−0.14	−0.25	−0.05	0.004	−0.42	0.69
AWHC	−0.36	0.06	−0.01	0.84	−0.05	0.84
P_{ava}	−0.11	0.76	−0.07	0.19	0.04	0.64
K_{ava}	−0.41	0.20	−0.09	−0.68	−0.24	0.74
TN	−0.14	0.79	0.04	−0.03	−0.02	0.68
FC	−0.16	0.03	−0.008	0.93	0.01	0.91
WSA	−0.10	0.01	0.94	−0.01	0.02	0.90
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	–

*Communality estimates.

variables, and when examined may suggest an underlying common factor that explains why these variables are correlated. Of the 14 possible factors, only the first five had eigenvalues > 1.0 (Table III). The factors with eigenvalue > 1 were retained, since eigenvalue < 1 indicated that the factor could explain less variance than could individual attributes (Shukla et al., 2006). The first and most important factor (Factor 1) explained 30.20% of the total variance. The second factor accounted for a further 19.71% of the total variance. Factors 1 through 5 collectively accounted for 77.80% of the total variance and inclusion of more factors did not significantly increase in total the explained variance (Figure 2).

Communalities of the 14 variables measured in the field indicated that these five factors explained a large part of the variation of most of the measured variables (Table III). More than 85% of the variation in MWD, FC, WSA, EC, CEC, and AWHC was

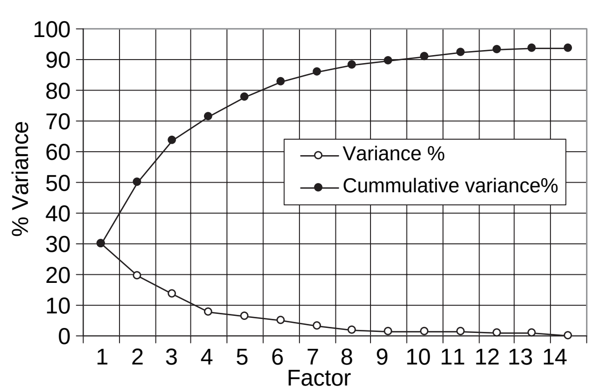

Figure 2. Relative and cumulative variances represented by 14 factors.

explained by the five factors. Measured variables with relatively high factor loadings within each factor correlation are indicated in Table III. Factor loadings indicate the correlation between a variable and an underlying common factor. These highly loaded variables were then used to propose a possible common underlying factor that linked variables together within each factor. The fact that two or more variables (Factors) are grouped in a latent variable suggests a possible common factor that makes them vary together within the field. The signs of the factor loadings provide information of how these variables relate when representing the common factor.

The interpretation of each latent variable (Factor) is an important aspect of factor analysis and no general rules can be provided (Mallarino et al., 1999). Agronomic knowledge of potential reasons for the observed co-variation and subjective judgment are involved. The underlying common factor represented by the Factors may not be readily obvious, but the results provide a basis for speculation. Factor 1 had the largest eigenvalue (Table III) and had the highest loadings (>0.6) for EC and SAR. It was termed the '*salinity and sodicity*' factor because EC and SAR showed the highest loadings in the first factor. Factor 2 was termed the '*fertility*' factor because of high positive loadings for P_{ava} and TN and high negative loading for pH. The high positive loadings for P_{ava} and TN verified the relatively high correlation between them ($r = 0.49$, $P < 0.01$, Table II). On the other hand the negative correlation between P_{ava} and pH suggested that availability of phosphorus strongly depended on the

Table IV. Coefficents, and multiple coefficient of determination (R^2), and root-mean-squared error (RMSE) for regression models for relating barley biomass with latent variables in the field of study (only factors significantly related to biomass included in the Table).

		Factors and latent variable name[a]					
Model	Intercept	Factor 1 *Salinity and sodicity*	Factor 2 *Fertility*	Factor 3 *Aggregation*	R^2	$P>F$	RMSE
1	2.53 (0.06)[b]	−0.82 (0.07)	–	–	0.60	0.001	0.080
2	2.53 (0.06)	−0.82 (0.06)	0.41 (0.06)	–	0.78	0.001	0.031
3	2.53 (0.06)	−0.82 (0.06)	0.41 (0.06)	0.15 (0.06)	0.79	0.001	0.047

[a]See text for interpretation of names assigned to these variables. [b]Numbers in parentheses are standard errors of the estimates.

soil acidity (Lindsay, 1979). Factor 3 was termed the '*aggregation*' factor because MWD and WSA had the highest positive loadings in this factor. These two variables were highly correlated ($r = 0.87$, Table II) to each other and with OM. Also, other authors (Islam & Weil, 2000; Celik, 2005) emphasized the influence of organic matter on soil aggregation. Factor 4 was termed the '*water availability*' factor because FC and AWHC showed the highest positive loadings (Table III). The positive coefficients of factor 2 (Table IV) showed the positive influence of soil fertility to improve barley biomass-yield in the field studied.

Regression analysis

Using factors as variables for multiple regression analysis (stepwise regression analysis) avoids the multicollinearity problems that are associated with multiple regression analysis using variables that are correlated with each other. Table IV shows coefficients and statistics of models relating barley biomass with the latent variables (Factors). The first factor explained almost 60% of the total variance and contrasted the negative influences of EC and SAR (Model 1, Table IV). The addition of one more parameter (Factor 2) to Model 1 increased the R^2 from 0.60 to 0.78 (Model 2, Table IV). The addition of one more factor (Factor 3) to the previous model increased R^2 from 0.78 to 0.79, which is not a pronounced improvement. The root-mean-square error (RMSE) of barley biomass prediction ranged from 0.08, through 0.031 to 0.047 for these three models, respectively. Because of the negligible increase in R^2 in Model 3 compared with Model 2, and the lowest RMSE and significant R^2 for Model 2, the selection of the 2-parameter model (Model 2, Table IV) was probably the best predictor for barley biomass. As a result, the first two factors were included in the regression equation. The most important term in the regression equation for barley biomass prediction was the salinity and sodicity, Factor 1, which had the largest coefficient. Interpretation of the signs of the coefficients of the models requires study of signs and relative weights of the factor loadings of the variable included in each latent variable (Table III) and simple correlations (Table II). For example, the negative sign of '*salinity and sodicity*' latent variable seems reasonable, because salinity restricts the plant production by different ways including: increased water stress, by reducing the osmotic potential, antagonistic behavior with nutrients, ion toxicity (Mass & Grieve, 1987; Corwin & Lesch, 2003; Khan et al., 2004); and sodicity destroys soil physical characteristics (Felhendler et al., 1974; Aggasi et al., 1981; Eltaif & Gharaibeh, 2007).This effect could have been overlooked by simple observation of a table of simple correlation, because both EC and SAR were correlated and showed significantly negative correlation with biomass and grain yield (Table II). These results are consistent with results reported by other authors, that a considerable variability in crop yields in salt-affected soils is explained by electrical conductivity (Kitchen et al., 2003).

Table V. Coefficents, and multiple coefficient of determination (R^2), and root-mean-squared error (RMSE) for regression models for relating grain yield of barley with latent variables in the field of study (only factors significantly related to grain yield included in the Table).

		Factors and latent variable name[a]					
Model	Intercept	Factor 1 *Salinity and sodicity*	Factor 2 *Fertility*	Factor 4 *Water availability*	R^2	$P>F$	RMSE
1	0.82 (0.02)[b]	−0.24 (0.02)	–	–	0.68	0.001	0.110
2	0.82 (0.02)	−0.24 (0.02)	0.10 (0.002)	–	0.72	0.001	0.066
3	0.82 (0.02)	−0.24 (0.02)	0.10 (0.002)	0.07 (0.002)	0.73	0.001	0.018

[a]See text for interpretation of names assigned to these variables. [b]Numbers in parentheses are standard errors of the estimates.

Data in Table V provide coefficients and statistics of models relating to grain yield with latent variables. The first factor explained 68% of the total variance of grain yield and contrasted the negative influences of EC and SAR (Model 1, Table V). Addition of two more factors (Factor 2 and Factor 4) increased R^2 from 0.68 to 0.72 and 0.73 for Model 2 and Model 3, respectively. The last model (Model 3) was identified as the best predictive model based upon R^2 and RMSE values, whereas this model showed the highest R^2 and the lowest RMSE (Table V). In this equation regression model Factor 3 (namely *aggregation*) was not significant. It means that, in the field studied, grain yield of barley was predominantly affected by *salinity and sodicity*, *soil fertility*, and *water availability* rather than by *soil aggregation*. It is important to note that despite having the same nomenclature (i.e., F_1, F_2 ...), the variables for each regression equation are independent of those used in other equations. Salinity and sodicity and soil fertility were identified as the most controlling factors for barley biomass and grain yield based upon the regression analysis. According to loadings of eigenvectors in significant factors, the most controlling soil properties were identified as EC, SAR, pH, TN, and P_{ava} for both biomass and grain yield. Moreover, AWHC and FC were determined as the soil properties which affected grain yield of barley in the study area. The significant difference between two developed models was related to water availability, which contributed to grain yield. It seems that water-retention capacity of soil had a valuable effect on grain-filling in the maturity stage of growth (Agueda et al., 2007).

The fact that several groups of correlated site variables could be identified for each field does not necessarily mean that they explain yield variability. The developed regression equations for prediction of barley biomass and grain yield explained 75 and 72% variance in biomass and grain yield. Therefore, the remaining variance may belong to nonmeasured variables like micronutrient variability and management effects such as irrigation pattern and weed control within the field. Furthermore, the prediction capability of these models, the latent variables that are significantly related to yield, can be useful to understand the reasons for biomass and grain-yield variability, and this understanding can, in turn, be used to manage better for crop production.

Soil variability expressed as CV was low for pH, BD, FC, K_{ava}, TN, and AWHC, moderate for SAR, CEC, WSA, and OM, and the highest for EC, MWD, P_{ava}, and CCE. Biomass and grain yield also had relatively high variability. Correlation analysis among soil variables showed that there are significant correlations in the great soil attributes pairs (>50%). Therefore factor analysis was established. Factor analysis provided a rational criterion for including and arranging correlated attributes in multiple regression models relating biomass and grain yield with soil attributes. The results showed that the first five factors had eigenvalues more than 1, and explained 77.80% of total variance. Each factor was termed based on loadings of soil variables in that factor. The first four factors which were significant in the regression models were termed as '*salinity and sodicity*' for factor 1, '*soil fertility*' for factor 2, '*aggregation*' for factor 3, and '*water availability*' for factor 4.

The stepwise regression identified factor 1, which explained 30.2% of variability, as the dominant indicator for predicting biomass and grain yields. The coefficients of factor 1 were negative for both grain and biomass yield. Therefore, improvement in soil drainage and application of amendments for rehabilitation of salinity and sodicity will increase biomass and grain yield of barley. Furthermore, the coefficient of factors 2 and 3 for biomass, and factors 2 and 4 for grain yield, were positive, indicating that improvement of soil fertility and soil attributes affecting water availability (such as soil structure and organic matter) are important to increase crop productivity. Overall, the results revealed that factor analysis was successful in identifying groups of correlated soil variables that were significantly correlated with the within-field yield variability. The results of this study are only applicable to the site studied and to other sites in the same region with similar topography, climate, soils, and management. However, the methodology used for analysing the data has wider applicability and can be applied to other sites and crops.

Acknowledgments

The authors gratefully appreciate Mr. Mojtaba Nouruzi for his assistance in fieldwork and laboratory. Comments on an earlier version of this paper by anonymous reviewers and the editor are greatly appreciated.

References

Adams, M.L., Zhao, F.J., McGrath, S.P., Nicholson, F.A., & Chambers, B.J. (2004). Predicting cadmium concentration in wheat and barley grain using soil properties. *Journal of Environmental Quality*, *33*, 532–541.

Agassi, M., Shainberg, I., & Morin, J. (1981). Effect of electrolyte concentration and soil sodicity on infiltration rate and crust formation. *Soil Science Society of America Journal*, *48*, 848–851.

Agueda, G., Isaura, M, & Luis, A. (2007). Response of barley genotypes to terminal soil moisture stress: phenology, growth, and yield. *Australian Journal of Agricultural Research*, *58*, 29–37.

Angers, D.A., & Mehuys, G.R. (1993). Aggregate stability to water. In: Carter, M.R. (Ed.), *Soil Sampling and Methods of Analysis* (pp. 651–657). Lewis Publishers, Boca Raton, FL, USA.

Ayoubi, Sh. (2005). Soil attributes prediction using soil-landscape model in hilly region of Azadshahr district, Golestan province, Northern Iran. Gorgan University of Agricultural Sciences and Natural Resources publication, GUASNR Press, Gorgan University, 54 pp.

Ayoubi, Sh., Mohammad Zamani, S., & Khormali, F. (2007). Spatial variability of some soil properties for site specific farming in northern Iran. *International Journal of Crop Production*, *1*, 225–236.

Blake, G.R., & Hartge, K.H. (1986). Bulk density. In: Klute, A. (Ed.), *Methods of Soil Analysis, Part I* (pp. 363–376), ASA, Monograph, No. 9. Madison, WI, USA.

Brejda, J.J., Moorman, T.B., Karlen, D.L., & Dao, T.H. (2000). Identification of regional soil quality factors and indicators: I. Central and southern high plains. *Soil Science Society of America Journal*, *64*, 2115–2124.

Bremner, J.M., & Mulvaney, C.S. (1982). Nitrogen total. In: Buxton, D.R. (Ed.), *Methods of Soil Analysis. Part II* (pp. 595–624), 2nd Edition, ASA, Monograph, No. 9, Madison, WI, USA.

Carr, P.M., Carlson, G.R., Jacoben, J.S., Nielsen, G.A., & Skogley, E.O. (1991). Farming soil, not fields: A strategy for increasing fertilizer profitability. *Journal of Production Agriculture*, *4*, 57–61.

Celik, I. (2005). Land use effects on organic matter and physical properties of soil in a southern Mediterranean highland of Turkey. *Soil & Tillage Research*, *83*, 270–277.

Corwin, D.L., & Lesch, S.M. (2003). Application of soil electrical conductivity to precision agriculture. *Agronomy Journal*, *95*, 455–471.

Cox, M.S., Gerard, P.D., Wardlaw, M.C., & Abshire, M.J. (2003). Variability of selected soil properties and their relationships with soybean yield. *Soil Science Society of America Journal*, *67*, 1296–1302.

Douaoui, A.E., Nicolas, K.H., & Walter, C. (2006). Detecting salinity hazards within a semiarid context by means of combining soil and remote sensing data. *Geoderma*, *134*, 217–230.

Eltaif, N.I., & Gharaibeh, M.A. (2007). Effects of single and mixed ion solutions on hydraulic and physical properties of a clay soil. *Water, Air & Soil Pollution*, *181*, 297–302.

Elustondo, J., Angers, D.A., Laverdiere, M.R., & N'dayegamiye, A. (1990). Influence de la culture de mais et de la prairie sur l'agrégation et la matiere organique de sept soils de Quebec. *Canadian Journal of Soil Science*, *70*, 395–403.

Felhendler, R., Shainberg, I., & Frenkel, H. (1974). Dispersion and hydraulic conductivity of soils in mixed solution. Transactions of the 10th International Congress of Soil Science, Moscow, *1*, 103–112. Moscow, Nauka Publishers.

Freund, R.J., & Littell, R.C. (2000). *SAS system for regression*. SAS Instiiute, Cary, NC, USA.

Godwin, R.J., & Miller, P.C.H. (2003). A review of the technologies for mapping within-field variability. *Biosystems Engineering*, *84*, 393–407.

Hair, J.F., Anderson, J.R.E., & Tatham, R.L. (1987). Multivariate data analysis with readings (2nd edn). Macmillan Publishing Co, New York, USA.

Islam, K.R., & Weil, R.R. (2000). Land use effects on soil quality in a tropical forest ecosystem of Bangladesh. *Agriculture, Ecosystem and Environment*, *79*, 9–16.

Jiang, P., & Thelen, K.D. (2004). Effects of soil and topographic properties on corn yield in a north-central corn-soybean cropping system. *Agronomy Journal*, *96*, 252–258.

Jolliffe, L.T. (1986). *Principal component analysis*. Springer-Verlag, New York, USA.

Kaspar, T.C., Fenton, D.J., Colvin, T.S, Karleno, D.L., Jaynes, D.B., & Meek, D.W. (2004). Relationships of corn and soybean yield to soil and terrain properties. *Agronomy Journal*, *96*, 700–709.

Kemper, W.D., & Rosenau, R.C. (1986). Aggregate stability and size distribution. In: Klute, A. (Ed.), *Methods of Soil Analysis, Part I* (pp. 687–734), ASA, 2nd edn, Monograph, No. 9, Madison, WI, USA.

Khakural, B.R., Robert, P.C. & Huggins, D.R. (1999). Variability of corn/soybean yield and soil landscape properties across a southwestern Minnesota landscape (pp. 573–579). In: Robert P.C. et al. (Ed.), *Proceedings of the Fourth International Conference on Precision Agriculture*, St. Paul, MN. 19–22 July 1998. ASA, CSSA, and SSSA, Madison, WI, USA.

Khan, A., Khan, A., & Nisar Ahmad, M.Z. (2004). Effect of salinity and sodicity on crop production at Kot Kashmir (Lakki Marwat). *Sarhad Journal of Agriculture*, *20*, 583–590.

Khormali, F., & Ayoubi, Sh. (2006). Clay mineralogy, micromorpholoy and physico-chemical analysis of the saline-sodic soils of GUASNR research farm, Aq Qala. Project Report, Gorgan University, 79 pp.

Kitchen, N.R., Drummond, S.T., Lund, E.D., Sudduth, K.A., & Buchleiter, G.W. (2003). Electrical conductivity and topography related to yield for three contrasting soil-crop systems. *Agronomy Journal*, *95*, 483–495.

Klute, A., & Dirksen, C. (1986). Hydraulic conductivity and diffusivity. In: Klute, A. (Ed.), *Methods of Soil Analysis, Part 1* (pp. 687–734), 2nd edn. Agronomy Monograph, Vol. 9. American Society of Agronomy, Madison, WI, USA.

Kovda, V.A. (1977). Arid land irrigation and soil fertility: problems of salinity, alkalinity, and compaction. In E.B. Worthington (Ed.), Arid Land Irrigation in Developing Countries: Environmental Problems and Effects (pp. 211–236). Pergamon Press, Oxford, UK.

Kravchenko, A.N., & Bullock, D.G. (2000). Correlation of corn soybean grain yield with topography and soil properties. *Agronomy Journal*, *92*, 75–83.

Lindsay, W.L. (1979). Chemical Equilibria in Soils. John Wiley & Sons Inc., NY, USA. 411 pp.

Lu, G., Sakagami, K., Tanaka, H., & Hamada, R. (1998). Role of soil organic matter in stabilization of water aggregates in soils under different types of land use. *Soil Science and Plant Nutrition*, *44*, 147–155.

Mallarino, A.P., Oyarzabal, E.S., & Hinz, P.N. (1999). Interpreting within field relationships between crop yields and soil and plant variables using factor analysis. *Precision Agriculture*, *1*, 15–25.

Mass, E.V., & Grieve, C.M. (1987). Crop salt tolerance current assessment. *Journal of Irrigation and Drainage Division*, *103*, 559–564.

Mclean, E.O. (1982). Soil pH and lime requirement. In: Page, A.L. (Ed.), Methods of Soil Analysis, Part II (pp. 199–223), 2nd edn. Monograph, No. 9, ASA, Madison, WI, USA.

Moore, I.D., Gessler, P.E., Nielsen, G.A., & Peterson, G.A. (1993). Soil attributes prediction using terrain analysis. *Soil Science Society of America Journal*, *57*, 443–452.

Nelson, D.W., & Sommers, L.E. (1982). Total carbon, organic carbon, and organic matter. In: Buxton, D.R. (Ed.), *Methods of Soil Analysis. Part II* (pp. 539–579), 2nd edn. Monograph, No. 9, ASA, Madison, WI, USA.

Olsen, S.R., & Sommers, L.E. (1982). Phosphorus. In: Page, A.L. (Ed.) Methods of soil analysis, Part 2: Chemical and microbiological properties (pp. 403–430), 2nd edn. American Society of Agronomy, Madison, WI, USA.

Ovalles, F.A., & Collins, M.E. (1988). Variability of northwest Florida soils by principal component analysis. *Soil Science Society of America Journal*, *52*, 1430–1435.

Page, M.C., Sparks, D.L., Noll, M., & Hendricks, G.J. (1987). Kinetics and mechanisms of potassium release from sandy middle Atlantic Coastal plain soils. *Soil Science Society of America Journal*, *51*, 1460–1465.

Quine, T.A., & Zhang, Y. (2002). An investigation of spatial variation in soil erosion, soil properties and crop production within an agricultural field in Devon. *U.K. Journal of Soil and Water Conservation*, *57*, 50–60.

Rhoades, J.D. (1982). Soluble salts. In: Page, A.L. (Ed.), *Methods of Soil Analysis, Part II* (pp. 167–179), 2nd edn. Monograph, No. 9, ASA, Madison, WI, USA.

Salinity Laboratory Staff (1954). Diagnosis and Improvement of Saline and Alkaline soils, USDA Agriculture Handbook No. 60. U.S. Government Printing Office, Washington, DC. 160 pp.

SAS Institute (2000). SAS Online Doc. Version 8 [Online]. Available at http://v8doc.sas.com/sashtml/ (accessed 2 Apr. 2003; verified 23 Feb. 2004). SAS Institute, Cary, NC, USA.

Shukla, M.K., Lal, R., & Ebinger, M. (2004). Principal component analysis for predicting corn biomass and grain yields. *Soil Science*, *169*, 215–224.

Shukla, M.K., Lal, R., & Ebinger, M. (2006). Determining soil quality indicators by factor analysis. *Soil & Tillage Research*, *87*, 194–204.

Soil Survey Staff (2006). Keys to Soil Taxonomy. U.S. Department of Agriculture, Natural Resources Conservation Service.

Swan, A.R.H., & Sandilands, M. (1995). *Introduction to Geological Data Analysis*. Blackwell, London. 446 pp.

van Bevel, C.H.M. (1949). Mean weight diameter of soil aggregate as a statistical index of aggregation. *Soil Science Society of America Proceedings*, *14*, 20–23.

van Es, H.M., Ogden, C.B., Hill, R.L., Schindelbeck, R.R., & Tsegaye, T. (1999). Integrated assessment of space, time, and management-related variability of soil hydraulic properties. *Soil Science Society of America Journal*, *63*, 1599–1608.

Vieira, S.R., & 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, T.R. (2005). Simulation analysis of relative yield advantage of barley and wheat in an eastern Mediterranean climate. *Field Crop Research*, *91*, 287–296.

Wang, Z., Chang, A.C., Wu, L., & Crowley, D. (2003). Assessing the soil quality of long-term reclaimed wastewater-irrigated cropland. *Geoderma*, *114*, 261–278.

Wendroth, O., Reynold, W.D, Vieira, S.R., Reichardt, K., & Wirth, S. (1997). Statistical approaches to the analysis of soil quality data. In E.G. Gregorich, & M.R. Carter (Eds.), *Soil Quality for Crop production and Ecosystem Health* (pp. 247–276). Elsevier, Amsterdam, The Netherlands.

Wilding, L.P. (1985). Spatial variability: Its documentation, accommodation, and implication to soil surveys. In D.R. Nielsen, & J. Bouma (Eds.), *Soil Spatial Variability*. Pudoc, Wageningen, The Netherlands.

Yusefi, S., Wissal, M., Mahmoudi, H., Abdelly, C., & Gharsalli, M. (2007). Effect of salt on physiological responses of barley to iron deficiency. *Plant Physiology and Biochemistry*, *45*, 309–314.