Modeling for relationships between soil properties and yield components of wheat using multiple linear regression and structural equation modeling

Leila Nazmi

Ph.D. of Soil Science, Armenian State Agrarian University, Yerevan, Armenia No. 26.Bahar 2, Sarein Station, Imam st. Ardabil, Iran

ABSTRACT

Since soil properties influence the behavior of soils, the knowledge related to these properties is important in using them for different agricultural purposes. This study aimed to develop a structural equation model of yield components of wheat (YCW) in northwest of Iran using soil physical and chemical properties. Soil samples were collected from Mollaahmad watershed of Ardabil province in northwest of Iran for the greenhouse experiment. The primary purpose of this research was to develop a conceptual model in order to determine the sources of variations within the dataset and to explore equations for the sampled soils. The findings revealed that two soil properties components (chemical and physical properties) were significant in explaining YCW. The accepted model in the multiple linear regression (MLR) analysis demonstrated that the soil's chemical and physical properties measures are statistically significant in estimating YCW. Following this, and according to R square statistic, 87% of the variance in YCW was explained by the soil chemical properties and 83% was accounted for by soil physical properties. Considering the relative importance of the estimation of YCW variable, and from the perspective of regression equations, the organic carbon and saturated point moisture made the largest contribution through the two proposed models for the soil productivity. According to the structural equation modeling (SEM) results, the final model has proved that YCW was controlled by soil chemical properties more than physical properties. The obtained general model can be useful for wheat and also an analytical pattern for Gramineae family. The improved estimation of production might be valuable in practice because crop productions are widely applied, for instance, to assess agro-environmental policy measures or the need of a better soil quality managing.

Key words: Crop yield, soil properties, multiple linear regression, structural equation modeling

Introduction

Several factors related to soil fertility including soil type, farmer’s practices, crop residues and mineral fertilizers management influence soil productivity and agricultural production [5, 38]. Soil productivity can be defined as the capacity of a soil, in its normal environment, to support plant growth such as wheat, rice and barley which relates directly to crop and forage yields and ties directly back to the portion of the soil fertility or quality definitions [15]. Crop production involves a complex interaction between the environment, soil parameters, and nutrient dynamics. Because of this fact, the soil must be studied in terms of the productive potentials. Failure to understand these complexities has resulted in lack of good crop production and management techniques; hence agricultural production has tended to be low [33]. Assessing soil fertility decline including nutrient depletion, nutrient mining, acidification (decline in pH and or an increase in exchangeable Al), loss of organic matter and increase in toxic elements (e.g., Al, Mn) is difficult because most soil chemical properties either change very slowly or have large seasonal fluctuations [21]. Thereby, a variety of mathematical models relating to crop yield have been recently proposed to evaluate causal relations between variables of soil properties and yield [27, 43, 20]. The present research also attempted to establish a mathematical yield model at the semi-arid region of northwestern Iran by incorporating the potential interactions between soils properties.

Using data obtained from experiments, MLR was used for evaluating equation linking several independent or predictor variables (soil chemical and physical properties) and one dependent variable (YCW). MLR based on soil chemical and physical properties, was given as an accurate tool to evaluate YCW, since it generated a minimum data set of...
indicators. Then, all the variables would be included simultaneously into single model in order to test the potential interactions between the independents variables using the structural equation modeling (SEM). In sum, the main purposes of this study were:

1. To examine the effects of soil chemical and physical properties on YCW.
2. To identify the most important parameters of soil physical and chemical properties in yield.
3. To suggest general model for strengthening soil productivity.

Materials And Methods

The study area is located in the Mollaahmad watershed of the Ardabil province in northwestern Iran (latitude: 38°03’23” to 38°07’46”N, and longitude: 48°10’58” to 48°21’13”E), with a mean annual temperature and precipitation of 19.3 °C and 303.5 mm, respectively. In this study, a landscape with the slope gradient of 1.4’ (flat area), uniform slope shape, slope direction of northwest and 120 meter slope length was selected. Four disturbed and undisturbed (by cylinder) soil samples with five replications were randomly collected from the soil surface (25-30 cm) for chemical and physical properties analysis. The properties measured were pH, electrical conductivity (EC) by saturated extracts, organic carbon content (OC) by potassium dichromate reduction of organic carbon and subsequent spectrophotometric measurement or with modified Walkley-Black method, total nitrogen (TN) by the wet oxidation of soil organic matter and botanical materials using a micro Kjeldahl procedure with sulfuric acid and digestion catalyst, available phosphorous (AP) by alkaline extraction by 0.5 normal NaHCO3, available potassium (AK) by ammonium acetate, 1 N, at pH=7.0 and subsequent determination by atomic absorption/emission spectrometry, calcium carbonate equivalent (CCE)) by gravimetric determination by reaction with hydrochloric acid, soil water content at saturation percentage (SP) by saturated soil paste, mean weight diameter (MWD) and geometric mean diameter (GMD) of soil aggregates by wet sieve in disturbed soil samples and bulk density (Pb) by measurement of volume and mass, saturated hydraulic conductivity (Ks) by the falling head method, field capacity (FC) and permanent wilting point (PWP) with pressure plate method in undisturbed soil samples. The soil properties were determined according to the methods described in Klute [25] and Sparks [40]. Along the same line, four composite soil samples with five replications were randomly collected for planting a spring wheat cultivar (Yan 7578128//Chil/2*Star) with the density of four plants per pots of 5 kg capacity. Overall, 20 pots were used in the greenhouse. N-P-K fertilizer including urea (130 kg ha⁻¹ equals 0.2 g pot⁻¹), super phosphate triple (150 kg ha⁻¹ equals 0.231 g pot⁻¹) and potassium sulphate (250 kg ha⁻¹ equals 0.385 g pot⁻¹) were used in the experimental pots. Irrigation in several stages and at normal times necessary for wheat growth took place. At the end of growth period, dry matter, grain yield, spike number, spike weight and 1000-grain weight were determined. In order to form relationship between soil physical and chemical properties and YCW, MLR analysis was performed by the LISREL 8.50 statistical program [6], at the same time the hypothesized structure was tested statistically using LISREL 8.5 software [23].

Development of Conceptual Model:

The improvement of the conceptual model was based on what is already acquired by the earlier research, that is to say, the realization of relationships and correlations among variables have already been validated and are already known. Owing to the role and importance of soil properties on crop production, the necessity for evaluation of soil properties to predict yield based on an integrated and efficient model is urgently required. The most of the earlier studies carried out in this respect utilize regression models [2,22] and SEM analysis [9] to show the relationships between soil properties and crop yield [46].

Therefore, for the realization of conceptual model of soil properties and YCW, some characteristics of soil variables on which based the present study were remembered briefly. There are many reports about the effects of N on wheat quality, yield and nitrate leaching, mainly from central Europe [1]. Nitrogen is the main and most frequently yield-limiting nutrient for high yields of most field crops [46,16]. In Lamb et al., [26] research the implicit effect of precipitation on seed number appears to be the main driver for final yield. Katan [24] declared that higher soil moisture resulted in seedling emergence. Decreasing plant available water capacity is the most serious loss associated with the long term productivity [28,33,34,42]. The key to raise crop yield, lies to a large extent, in the increase of usable water and raising the efficiency of water use [29]. According to the results of other researches, nitrogen and phosphorus are the most limiting nutrients to crop production [31]. Soil factors that affected yields most were soil moisture-holding capacity, cation exchange capacity (CEC), soil texture, pH, extractable P, and percent base saturation. The particle size distribution of the plot with high sand texture (66.20 and 71.80%) and lower silt texture (7.20 and 9.30%) improved significantly values of dry weight aerial parts, roots dry weight, number of flowering per plant and flowering rate. Zhao et al., [46] claimed that soil EC could serve as a guideline for soil productive potentiality, and thus be
used as an important indicator to assess the variability of crop yield. Soil bulk density reduction which resulted in better soil aeration and improved soil structure, increase the crop production [19]. An increase of FC and PWP due to the high organic matter content, increases the soil productivity [7].

We considered the use of MLR as a statistical technique for modeling multiple observed variables [39]. The MLR of theoretical interest in current paper was to predict YCW based on parameters of soil properties. Then, all the variables would be included simultaneously into single model in order to test the potential interactions between the independents and dependent variables using the SEM [30]. The integrated result provided a method to characterize YCW and soil physical properties. We aimed to prove the following: (H1) that the latent variables of soil chemical and (H2) soil physical properties that had a positive impact on YCW.

Results And Discussion

Multiple Linear Regression (MLR):

The theoretical regression model included a set of seven independent explanatory variables for each of the soil chemical and physical properties, separately. MLR method provided equation linking dependent variable Yd (YCW) to the independent variable Vi using the following form:

\[ Yd = \beta_0 + \beta_1 V_{i1} + \ldots + \beta_n V_{in} \]

When the intercept \( \beta_0 \) and the regression coefficients of descriptors (\( \beta_i \)) are determined by leastsquares method (Green and Carroll, 1996). Vi descriptors are used to describe soil chemical and physical properties. The reduction in the number of descriptors (variables) is included in the study to minimize the information overlap in variables. The best equation is selected while being based on the highest multiple correlation coefficients (R), lowest standard deviation (SD) and F-ratio value [11].

Regression analysis (LISREL 8.50) was conducted to investigate the relationships between YCW and soil chemical properties. The [pH], [EC], [CCE], [SO4], [OC], [TN], [AP] and [AK] were considered as independent variables and YCW as a dependent variable. The best model was derived by the application of MLR method. The descriptors and the regression coefficient of this model are presented in Table 1. As can be seen in the case of all the MLR analysis, the soil chemical properties measures were statistically significant in estimating YCW (P < 0.001).

The multiple R coefficient indicated that the correlation between soil chemical properties and YCW was moderate (the multiple R = 0.77), according to R square statistic, 75 % for the total variance for the estimation of YCW was explained by the MLR model. In terms of the relative importance of the estimation of a dependent variable, it can squabble that the [OC] made the largest contribution across the model. An examination of t values also revealed an identical descending order of the factors that contributed to the estimation of YCW.

Figure 1 presents diagrammatic representations of a MLR in which pH, EC, CCE, OC, TN, AP, and AK were related to YCW and 75% of the YCW level score variance was explained by them. The selected equation for soil chemical properties with standardized coefficients was:

\[ YCW=0.33*pH+0.25*EC-0.18*CCE+0.46*OC+0.14*TN+0.26*AP+0.12*AK \]

Therefore, the accepted model (Fig. 1) explained a moderate amount of the variance in YCW \( (R^2 = 0.75) \) by soil chemical properties. The regression analysis performed to evaluate the wheat yield model resulted in equation 1 that soil chemical properties prediction was significant for YCW \( (P<0.001, R^2 = 0.75) \). In terms of the relative importance of the estimation of a dependent variable, it can squabble that the OC made the largest contribution across the model (path coefficient=0.46). The positive sign of the beta coefficients and t-values pertaining to these variables indicates that there was a positive relationship between YCW and soil chemical properties ([pH], [EC], [OC], [TN], [AP],[AK]), and CCE negatively correlated to YCW. If the soil chemical properties including pH, EC, OC, TN, AP, AK increase, and CCE decrease, the amount of YCW will increase.

As can be seen in this equation, organic carbon had the highest effect on wheat yield in the study area. Productivity and yield parameters were not significantly influenced by nitrogen, phosphorus and potassium. The soil organic matter played an important role in maintaining physical, chemical and biological properties of the soil and the crop yield. Greater organic matter contents have been linked to increased water retention capacity in soils, especially at field capacity water content. This is believed to be caused by enhanced aggregate formation resulting from organic substances which increase the soil productivity. The negative effect of OC reduction on agricultural productions has been reported by previous researchers [35,45]. Crop yields in large parts of Kenya had been low due to declining soil fertility as a result of soils impoverishment in organic matter content and low reserves of nitrogen, phosphorus and some trace elements [4]. Therefore, enhancement and maintenance of soil productivity through the improvement of soil organic matter is one of the essential aspects for sustained agricultural production systems in the northwest of Iran.
Figure 2 depicts the standardized coefficients by soil physical properties on YCW. The regression coefficients for the paths in model represented 68% of the variance in YCW was accounted for by soil physical properties. The regression analysis, however, indicated that the standardized regression weight for all variables were statistically significant. The obtained results of standardized coefficients in regression model showed that the prediction equation for YCW was formulated using the soil physical properties variables as follows:

\[ YCW = 0.31 \times SP + 0.12 \times MWD + 0.08 \times GMD + 0.09 \times Ks - 0.11 \times Pb + 0.16 \times FC + 0.08 \times PWP \]  

This equation showed that soil physical properties prediction was also significant for YCW (\( \rho < 0.001 \), \( R^2 = 0.66 \)). Wheat yield was positively correlated to MWD, GMD, SP, FC, PWP, Ks and negatively correlated to soil bulk density. Soil moisture content allocated the highest contribution in determining the crop yield in this research (path coefficient=0.31). Similar results were noted by Nazmi et al., (2011), based on SP dramatic influence on wheat yield in the cultivated steepy lands.

**Hypothesis Testing By SEM:**

The third step of this study included simultaneously all the variables in a conceptual model in order to test the potential interaction between them. The SEM was used to achieve the goal and standardized coefficients were used to evaluate the strength of path coefficients estimated.

Figure 3 show that two hypotheses indicated causal relationships between soil (1) chemical and (2) physical properties with YCW. Support was found for hypothesis H₁ and H₂ approached significance (\( \hat{b} = 0.72, \ 0.64, \ p = 0.05 \)) respectively, therefore research hypotheses were approved and soil properties had significant positive effects on YCW.

Table 2 indicated that the theoretical model adequately constructed the observed covariance of the data with all of the aforementioned indicators of good model fit exceeding the minimum specifications. The \( \chi^2 \) statistic was non-significant, indicating a good model fit to the sample variance-covariance matrix (\( \chi^2 = 289.42, \ d.f = 236, \ \chi^2/df = 1.23, \ p = 0.45 \)) and descriptively (GFI = 0.93, RMSEA = 0.04).
Fig. 2: Regression model of YCW by soil physical properties

Fig. 3: Accepted theoretical model
All factors loadings that were tested had t-values greater than 1.96 and all of the path coefficients were significant. The goodness of fit indices for the structural model that are shown in Table 2 indicated the model has a good fit of the data. The chi-square ($x^2$) test yielded a value of 289.42, which had a corresponding p-value of 0.45, meaning that the chi-square/degrees of freedom was below 2 ($df=236$, $x^2/df=1.23$) normally a ratio in the range of 2–1 or 3–1, was indicative of an acceptable fit (Cote et al. 2001) and also fitted the data well, with a root mean square error of approximation (RMSEA) of 0.05 and a goodness-of-fit index (GFI) value of 0.93. These index values showed that there was a strong correspondence between predicted and observed covariances. For the RMSEA, a p value greater than 0.05 indicated no significant deviation between expected and observed covariances. For the GFI, values of 0.90 or more generally correspond to a high degree of fit for the structural model. Indicators generally showed a strong correlation to their respective latent variables. Several authors [44,17] mostly explained negative or positive yield trends due to several long-term management practices with trends in soil chemical properties (SOC, available N, P and K) or available nutrient status. Brahim et al., [9] declared that SEM modeling results show that model has proved that organic carbon was controlled by chemical properties more than physical properties. Also, De la Rosa et al., [14] reported that soil chemical properties including carbonate content, salinity, sodium saturation, and cation exchange capacity contributed to 78% of the variation in winter wheat yields. Ossom and Rhykerd [36] found that soil chemical properties influenced plant growth, and development, as well as the concentrations of various mineral nutrients at the end of the cropping season.

### Table 1: Summary of the linear regression predicting YCW

<table>
<thead>
<tr>
<th>Soil properties</th>
<th>$\beta$</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical properties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ph</td>
<td>0.33</td>
<td>12.45</td>
</tr>
<tr>
<td>EC</td>
<td>0.25</td>
<td>5.64</td>
</tr>
<tr>
<td>CCE</td>
<td>-0.18</td>
<td>7.28</td>
</tr>
<tr>
<td>OC</td>
<td>0.46</td>
<td>14.71</td>
</tr>
<tr>
<td>TN</td>
<td>0.14</td>
<td>6.22</td>
</tr>
<tr>
<td>AP</td>
<td>0.26</td>
<td>9.65</td>
</tr>
<tr>
<td>AK</td>
<td>0.12</td>
<td>4.88</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Multiple R</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>Physical properties</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SP</td>
<td>0.31</td>
<td>9.12</td>
</tr>
<tr>
<td>MWD</td>
<td>0.12</td>
<td>6.54</td>
</tr>
<tr>
<td>GMD</td>
<td>0.08</td>
<td>3.18</td>
</tr>
<tr>
<td>Ks</td>
<td>0.09</td>
<td>5.28</td>
</tr>
<tr>
<td>Pb</td>
<td>-0.11</td>
<td>5.82</td>
</tr>
<tr>
<td>FC</td>
<td>0.16</td>
<td>8.33</td>
</tr>
<tr>
<td>PWP</td>
<td>0.08</td>
<td>4.41</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>Multiple R</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Standard error</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Results of structural model tests—goodness of fit summary

<table>
<thead>
<tr>
<th>Measure</th>
<th>Full model values</th>
<th>Standard for acceptance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square</td>
<td>289.42</td>
<td>NA</td>
</tr>
<tr>
<td>$d_f$</td>
<td>236</td>
<td>NA</td>
</tr>
<tr>
<td>p-value</td>
<td>0.45</td>
<td>&gt; 0.05</td>
</tr>
<tr>
<td>$X^2/df$</td>
<td>1.73</td>
<td>&lt; 2</td>
</tr>
<tr>
<td>GFI</td>
<td>0.93</td>
<td>&gt; 0.90</td>
</tr>
<tr>
<td>RMSEA*</td>
<td>0.04</td>
<td>&lt; 0.05</td>
</tr>
</tbody>
</table>

*Goodness of Fit Index
*Root Mean Square Error of Approximation

**Conclusion:**

This study demonstrated how MLR could be effectively used for soil properties and to predict wheat yield. The MLR analysis showed that the soil chemical properties were the most important parameters in determining wheat yield variability; the OC and SP represented the key variables responsible for wheat yield. This statistical method can be used to predict yield for other crops and can be extended to other lands as well, where crop production is primarily dependent on soil properties.
Therefore, the model developed in MLR can be useful for forecasting crop production in regional and global scale if be conducted in future with the use of long period dataset.

Moreover, SEM provided an adequate explanation for the simultaneously interaction among the variables included in the conceptual model. SEM illustrated that the soil chemical factors were better indicators of wheat yield than what the physical properties did and YCW was dominated by OC and SP results of regression equations.

The integration of MLR and SEM analysis to evaluate YCW showed that crop yield was related to soil properties and YCW interaction. It appears that the methodology illustrated in this paper allows incorporating soil properties information (variables) into a statistical model and takes all possible interactions among the variables into consideration. Moreover, the proposed methodology can be useful in crop yield where soil properties data are available.

References


