



**Development of a Multivariate Regression Model for Soil Nitrate Nitrogen Content Prediction\***

*Xixi Wang, Assefa M. Melesse, and Wanhong Yang<sup>1</sup>*

**ABSTRACT**

Although soil nitrate nitrogen (N) is a nutrient source for crop, it could be a potential nonpoint pollution source to the environment when its content remains high with an inappropriate management. Soil nitrate N content is affected by various factors, such as cultivation practices, N fertilizer application rate, soil properties, and climatic conditions. Understanding the effects of these factors on soil nitrate N content is necessary for nitrogen management and nonpoint source pollution control. Taking the data measured from 1996 to 1998 in a 25 ha row crop field located in Central Iowa, this paper intended to study the interwoven effects of these factors on soil nitrate N content using multivariate statistical analysis techniques of sample mean plots, a multivariate analysis of variance (MANOVA) model, and a multivariate linear regression model. The inferences made by the sample mean plots and MANOVA model indicate that the effects of these factors are additive, i.e., their main or direct effects are statistically significant but the interaction effects between and among them are insignificant at a 5% significance level. Incorporating these additive effects, a multivariate linear regression model was fitted to the dataset. The residual plots show that the dataset follows an approximate bivariate normal distribution, which is assumed by the MANOVA and multivariate linear regression models. The validation using the field data collected in 1999 indicated that the model explained more than 93% variations exhibited by the measured sublayer-averaged data on soil nitrate N content and soil moisture. However, this model is unable to account for the within-sublayer variations.

**Key words:** linear regression; multivariate statistics; nitrogen; NPS; soil moisture; spatial analysis; visualization.

**INTRODUCTION**

Crop uptakes from root zone the needed nitrogen (N) that has an inorganic format (Gentry et al., 1998). With an inorganic structure and a high solubility, nitrate is a major nitrogen source for crop. Researchers (e.g., Sander et al., 1994; Olness et al., 1997, 1998; Algerbo and Thylen, 1998; Selles and James, 1998) found that maintaining soil nitrate N content to an appropriate value can achieve a maximal crop yield. A soil nitrate N content lower than this value may result in a reduced crop yield, whereas, a content higher than this value tends to increase the risk of environmental pollution because the residual nitrate N could be transported into water bodies or leached into groundwater (Baker and Lafen, 1982; Chaney, 1990; CAST, 1991; Lykins and Clark, 1995; Marshall and Bennett, 1998).

---

\*accepted March 2006.

<sup>1</sup> *Respectively, Ph.D., P.E., Research Scientist, Energy & Environmental Research Center, University of North Dakota, 15 North 23rd Street, Grand Forks, ND 58201; Assistant Professor, P.E., Department of Environmental Studies, Florida International University, Miami, Florida 33199; and Associate Professor, Department of Geography, University of Guelph, Guelph, Ontario N1G 2W1, Canada. (E-mail/Wang: xxqqwang@gmail.com).*

Soil nitrate N content is affected by various factors. It could be drastically changed by applying N fertilizer (Gentry et al., 1998). Albeit, soil nitrate N content may not be linearly increased as a result of the increase of N fertilizer application rate (Chaney, 1990). Microorganisms may adsorb and incorporate fertilizer N into their biomass before a crop's root system is sufficiently developed, which immobilizes the inorganic nitrogen until carbon becomes limiting as regulated by soil moisture (Gentry et al., 1998). Using data from 12 laboratory incubation studies for a total of 33 different soils, Paul et al. (2003) found a positive relationship between net nitrogen mineralization and soil moisture. In a separate study, Klodivko and Keeney (1987) concluded that this relationship is not noticeably affected by soil temperature but dependent upon soil properties.

In addition, soil nitrate N content is also related to climatic conditions (e.g., solar radiation and precipitation) and affected by other agricultural practices (e.g., tillage and rotation). A dry growing season, which limits crop growth and thus N uptake, could result in a high soil nitrate N content. Further, contrasting crop rotations, such as corn–soybean–corn versus soybean–corn–soybean, could result in distinctly different nitrate N contents (Sander et al., 1994; Gentry et al., 1998). A field with a soybean–corn–soybean rotation is expected to have a higher soil nitrate N content than that with a corn–soybean–corn rotation by the end of the third planting year. However, the difference resulting from crop rotation likely becomes undetectable, depending upon affects of other factors such as soil properties and precipitation (Drury et al., 1996).

In the literature, the relationship between soil nitrate N content and its affecting factors was primarily studied using laboratory experiments (e.g., Klodivko and Keeney, 1987; Paul et al., 2003) or plot tests (e.g., Chaney, 1990; Richards et al., 1995). While fundamentally important, the results from these studies might have limitations in practice. In addition, few studies examined whether and how the factors interactively affect soil nitrate N content. The objectives of this study were to 1) explore how soil nitrate N content is affected by various factors and their interactions for a 25 ha row-crop field located in Central Iowa, and 2) develop a relationship applicable for predicting soil nitrate N content at field scale.

The factors considered in this study included N fertilizer application rate, soil properties, soil moisture, climate, and agricultural practices. Limited by the small number of sampling points (less than 40) across the field, this study could not consider the spatial variability of these factors. Instead, the data collected at these sampling points were lumped into one analysis dataset. That is, the soil for a given horizon was assumed to be homogeneous across the field, while the soils for different horizons were considered to have distinctly different properties. In addition, to simplify the analysis, this study grouped these factors into three classes of: 1) N fertilizer application rate, designated Factor I; 2) soil depth, designated Factor II, which was used to represent the vertical variations of soil properties; and 3) year, designated Factor III, which was used to represent soil moisture, climate, and agricultural practices such as crop rotation. With this simplification, affects of cultivation practices on soil nitrate N content are indirectly reflected by Factor III.

## MATERIALS AND METHODS

### ***Field Measurement and Data Collapsing***

In the study field, corn was planted in 1996 and 1998 and soybean in 1997. For each of these three years, the field was subdivided into six to eight strips, which received N fertilizer treatments at low, medium, and high rates (Bakhsh et al., 2000). Data on soil nitrate N content and soil moisture were measured at randomly selected points after harvest. The measurement was implemented at 40 points in 1996 and nine points both in 1997 and 1998. For a given point, data were collected at eight depth intervals, designated Sublayer 1 to 8 for description purposes, which are 0–6, 6–12, 12–18, 18–24, 24–30, 30–36, 36–42, and 42–48 cm. For each sublayer, one to three soil samples were randomly taken and analyzed, formulating a dataset with a sample size of 962 for the three years.

In addition, a rainfall gauging station was set up to record the precipitation in the field (Bakhsh et al., 2000). An analysis of the historical precipitation data collected at Ames Municipal Airport, which is located about 60 km north of the study field, indicated that 1996, 1997, and 1998 were dry, wet, and normal years, respectively. Thus, these three years were expected to represent a wide range of climatic conditions to be experienced by the field.

As mentioned above, Factor I refers to N fertilizer application rate, Factor II to soil depth, and Factor III to year. Using these three classes of factors, a three-way contingency table (Neter et al., 1996) was formulated for the measured data on soil nitrate N content and soil moisture (not shown due to the large size). Again, Factor II reflects soil properties that are vertically varied, whereas, Factor III represents affects of climate and cultivation practices including crop rotation.

In 1999, soybean was planted and the field received no N fertilizer treatment. This year was a normal year in terms of the recorded precipitation. After harvest, data on soil nitrate and soil moisture were measured at nine points for the aforementioned eight depth intervals. These data were used for model validation.

### ***Analysis Method***

For the three-way contingency table, the sample means of soil nitrate N content and soil moisture were plotted to visually identify the significant main effects of, and interactions among, these three factors (Neter et al., 1996). In addition, a three-way multivariate analysis of variance (MANOVA) model (Johnson et al., 1999) was used to test findings from the visual plots. To minimize influences of uneven sample sizes, a harmonic mean of the sample variances was used in this test (Neter et al., 1996). The MANOVA model can be written as:

$$\begin{cases}
 \underline{y}_{ijkl} = \underline{\mu} + \underline{\tau}_i + \underline{\beta}_j + \underline{\gamma}_k + (\underline{\tau\beta})_{ij} + (\underline{\tau\gamma})_{ik} + (\underline{\beta\gamma})_{jk} + (\underline{\tau\beta\gamma})_{ijk} + \underline{\varepsilon}_{ijkl} \\
 \sum_{i=1}^3 \underline{\tau}_i = \sum_{j=1}^8 \underline{\beta}_j = \sum_{k=1}^3 \underline{\gamma}_k = \sum_{i=1}^3 (\underline{\tau\beta})_{ij} = \sum_{i=1}^3 (\underline{\tau\gamma})_{ik} = \sum_{j=1}^8 (\underline{\tau\beta})_{ij} = \sum_{k=1}^3 (\underline{\tau\gamma})_{ik} = \sum_{j=1}^8 (\underline{\beta\gamma})_{jk} = \sum_{k=1}^3 (\underline{\beta\gamma})_{jk} = \sum_{i=1}^3 (\underline{\tau\beta\gamma})_{ijk} = \sum_{j=1}^8 (\underline{\tau\beta\gamma})_{ijk} = \sum_{k=1}^3 (\underline{\tau\beta\gamma})_{ijk} \quad (1) \\
 i = 1, 2, 3 \\
 j = 1, 2, 3, \dots, 8 \\
 k = 1, 2, 3 \\
 l = 1, 2, \dots, n_{ijk}
 \end{cases}$$

where  $n_{ijk}$  – the sample size at  $i$ th level of Factor I,  $j$ th level of Factor II, and  $k$ th level of Factor III;  
 $\underline{y}_{ijkl}$  – the  $l$ th measurement vector ( $2 \times 1$ ) at  $i$ th level of Factor I,  $j$ th level of Factor II, and  $k$ th level of Factor III;  
 $\underline{\mu}$  – the overall level vector ( $2 \times 1$ );  
 $\underline{\tau}_i$  – the main effect vector ( $2 \times 1$ ) of Factor I;  
 $\underline{\beta}_j$  – the main effect vector ( $2 \times 1$ ) of Factor II;  
 $\underline{\gamma}_k$  – the main effect vector ( $2 \times 1$ ) of Factor III;  
 $(\underline{\tau\beta})_{ij}$  – the interaction effect vector ( $2 \times 1$ ) between Factor I and Factor II;  
 $(\underline{\tau\gamma})_{ik}$  – the interaction effect vector ( $2 \times 1$ ) between Factor I and Factor III;  
 $(\underline{\beta\gamma})_{jk}$  – the interaction effect vector ( $2 \times 1$ ) between Factor II and Factor III;  
 $(\underline{\tau\beta\gamma})_{ijk}$  – the interaction effect vector ( $2 \times 1$ ) among Factor I, Factor II, and Factor III;  
 $\underline{\varepsilon}_{ijkl}$  – the random error vector ( $2 \times 1$ ).

Model (1) assumes that  $\underline{\varepsilon}_{ijkl}$  has an independent bivariate normal distribution with a common covariance matrix (Johnson et al., 1999), which reflects the correlation between soil nitrate N content and soil moisture. Although this assumption could be relaxed in this study because of the large same size of 962, it was verified using residual plots (Neter et al., 1996). Subsequently, with independent variables representing the statistically significant effects and interactions identified using the sample mean plots and Model (1), a multivariate linear regression model was fitted to the dataset. The regression model was developed to exclude statistically insignificant variables to minimize the model prediction uncertainty. This model has a similar assumption with Model (1), which was also verified using residual plots. In addition, the regression model was validated using the 1999 field data. With this regard, the variable for N fertilizer application rate was assigned as low.

## RESULTS AND DISCUSSION

### ***The Sample Means of Soil Nitrate N Content***

The sample means of soil nitrate N content are shown in Figure 1. In these plots, the ordinate denotes soil nitrate N content, and the abscissa Sublayer. The sample means were computed from the three-way contingency table and plotted for the three study years of 1996, 1997, and 1998. Each of the plots consists of three lines, which correspond to the low, medium, and high N fertilizer application rates.

The main effect of soil depth (Factor II) is expected as indicated by that the lines in Figure 1 vary with Sublayer. Soil nitrate N content reached its peak in the first or second sublayer regardless of N fertilizer application rate and year. Thus, soil nitrate N content is highly correlated with soil depth, which is consistent with Gentry et al. (1998). In addition, for a given year, the main effect of N fertilizer application rate is expected.

However, the effect varied with climatic conditions and soil depth. In 1996 (dry year), the descending sequence of soil nitrate N content in Sublayer 1 corresponded to the medium, low, and high application rates, respectively, whereas, in Sublayers 2–5 to the low, medium, and high rates, respectively. In Sublayers 6–8, the descending sequence corresponded to the high, low, and medium rates, respectively. In 1997 (wet year), the descending sequence of soil nitrate N content in Sublayers 1–3 corresponded to the high, low, and medium application rates, respectively, whereas, in the other sublayers to the high, medium, and low rates, respectively. In 1998 (normal year), the descending sequence in Sublayers 1, 2, 7, and 8 corresponded to the high, medium, and low application rates, respectively, but in Sublayer 3 to the high, low, and medium rates, respectively, in Sublayers 4 and 5 to the low, medium, and high rates, respectively, and in Sublayer 6 to the medium, low, and high rates, respectively.

As a result of the positive and negative effects of N fertilizer application rate, the interaction effects among Factor I, II, and III became weak. For a given year, there was no interaction effect between N fertilizer application rate and soil depth, as indicated by that the three lines are almost parallel (Figure 1). In addition, for a given sublayer, the absolute differences of the soil nitrate N contents corresponding to the different application rates are almost equal, implying that no interaction effect between the application rate and year is expected. Further, for a given application rate, the three lines, which represent the soil nitrate N contents in 1996 (Figure 1a), 1997 (Figure 1b), and 1998 (Figure 1c), are almost parallel, indicating no interaction effect among the three factors. However, the soil nitrate N contents were different from year to year and from sublayer to sublayer, implying that the main effect of year (Factor III) is expected.

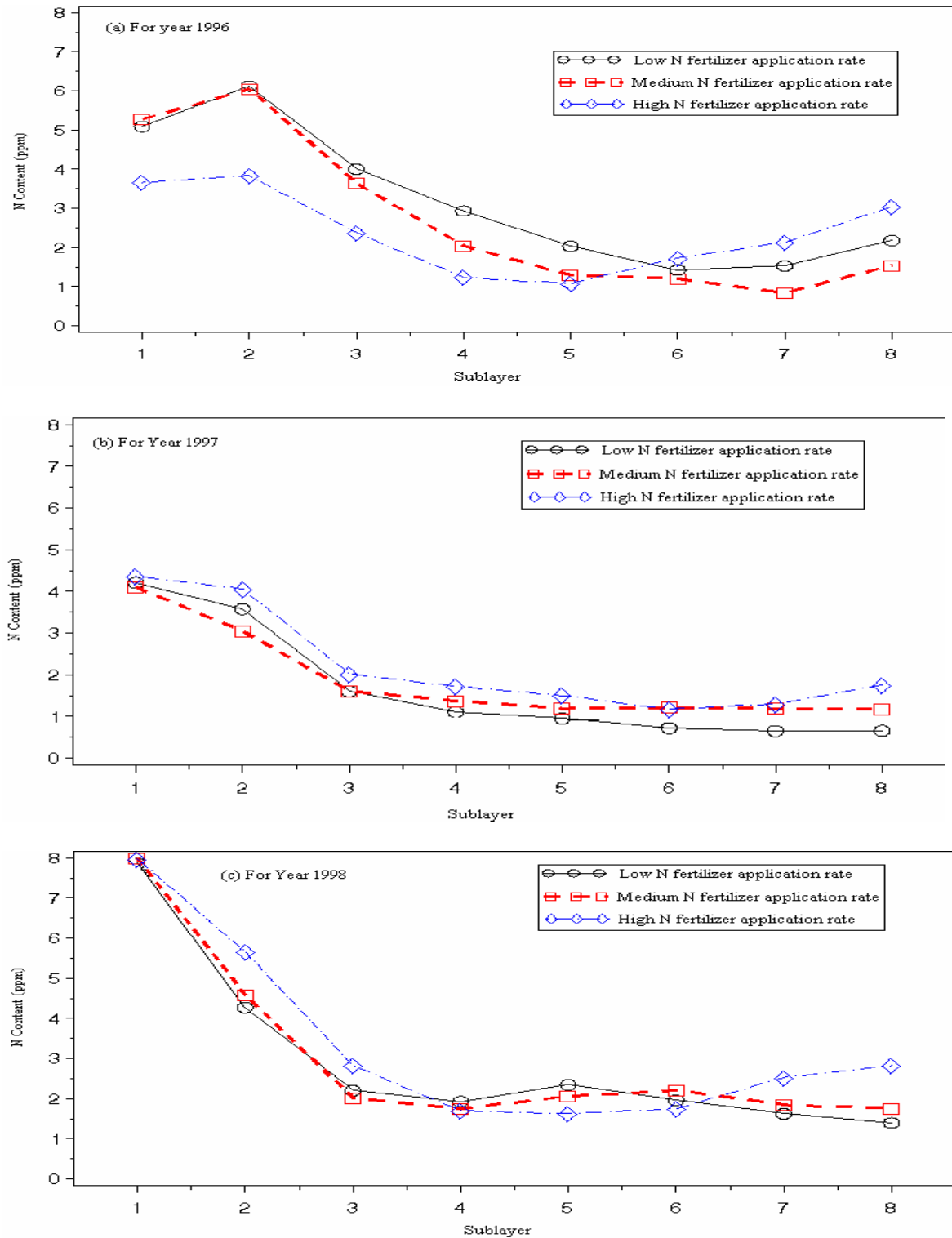


Figure 1. Sample mean plots of soil nitrate N content for the three years and N fertilizer application rates.

### **The Sample Means of Soil Moisture**

The sample means of soil moisture are shown in Figure 2. Similarly, in these plots, the ordinate denotes soil moisture, and the abscissa Sublayer. The sample means were computed from the three-way contingency table and plotted for the three study years of 1996, 1997, and 1998. Each of the plots consists of three lines, which correspond to the low, medium, and high N fertilizer application rates.

The main effect of soil depth (Factor II) on soil moisture is expected, as indicated by that the lines in Figure 2 vary with Sublayer. The overall pattern is that a lower sublayer tended to have a lower value for soil moisture than the upper sublayers. However, in 1997 and 1998, Sublayer 2 had a higher value for soil moisture than Sublayer 1. This might be because the crop probably used more water from Sublayer 1.

In addition, for a given year, the main effect of N fertilizer application rate on soil moisture is expected. In Sublayers 2–4, the descending sequence of soil moisture corresponded to the high, medium, and low application rates, respectively. A high N fertilizer application rate tends to prompt crop growth, resulting in a more developed crop root system (Jaynes et al., 1999). The enriched root system likely absorbed more water and thus lowered soil moisture. For the other sublayers, the factor of N fertilizer application rate either had a positive or negative effect on soil moisture, depending upon the climatic conditions. As with that on soil nitrate N content, the interaction effects among Factor I, II, and III became weak as a result of the positive and negative effects. For a given year, there was no interaction effect between N fertilizer application rate (Factor I) and soil depth (Factor II). In addition, for a given sublayer, the absolute differences of the values for soil moisture corresponding to the different N fertilizer application rates are almost equal, implying that no interaction effect between Factor I and Factor III (year) is expected. Further, for a given N fertilizer application rate, the three lines representing the soil moistures in 1996 (Figure 2a), 1997 (Figure 2b), and 1998 (Figure 2c) are almost parallel, indicating that the three factors did not interactively affect soil moisture. However, the soil moistures were different from year to year and from sublayer to sublayer, implying that the main effects of these three factors are expected.

### **The MANOVA Model**

The above inferences based on the sample mean plots indicate that the three factors independently rather than interactively affected soil nitrate N content and soil moisture. These inferences could be further refined by using MANOVA Model (1), which considers the variance of, and the correlation between, soil nitrate N content and soil moisture. The statistics used to make inferences, namely Wilks' Lambda, F value, and p value, were computed using the SAS software package (SAS Institute Inc., 1996) and are summarized in Table 1. The p values corresponding to the main effects of N fertilizer application rate, soil depth, and year are 0.0184, 0.0001, and 0.0001, respectively, indicating that the main effects are statistically significant at a 5% significance level. However, neither the two-factor interaction effects nor the three-factor interaction effect were determined to be significant (p values > 0.05). Thus, the inferences based on Model (1) are consistent with that based on the sample mean plots. To verify the MANOVA model's assumption of bivariate normality, the residuals were plotted versus the predicted soil nitrate N content and soil moisture (Figure 3). These plots exhibit no

obvious pattern. That is, the residuals do not consistently increase or decrease with the predicted values, indicating that the assumption of Model (1) holds for the dataset.

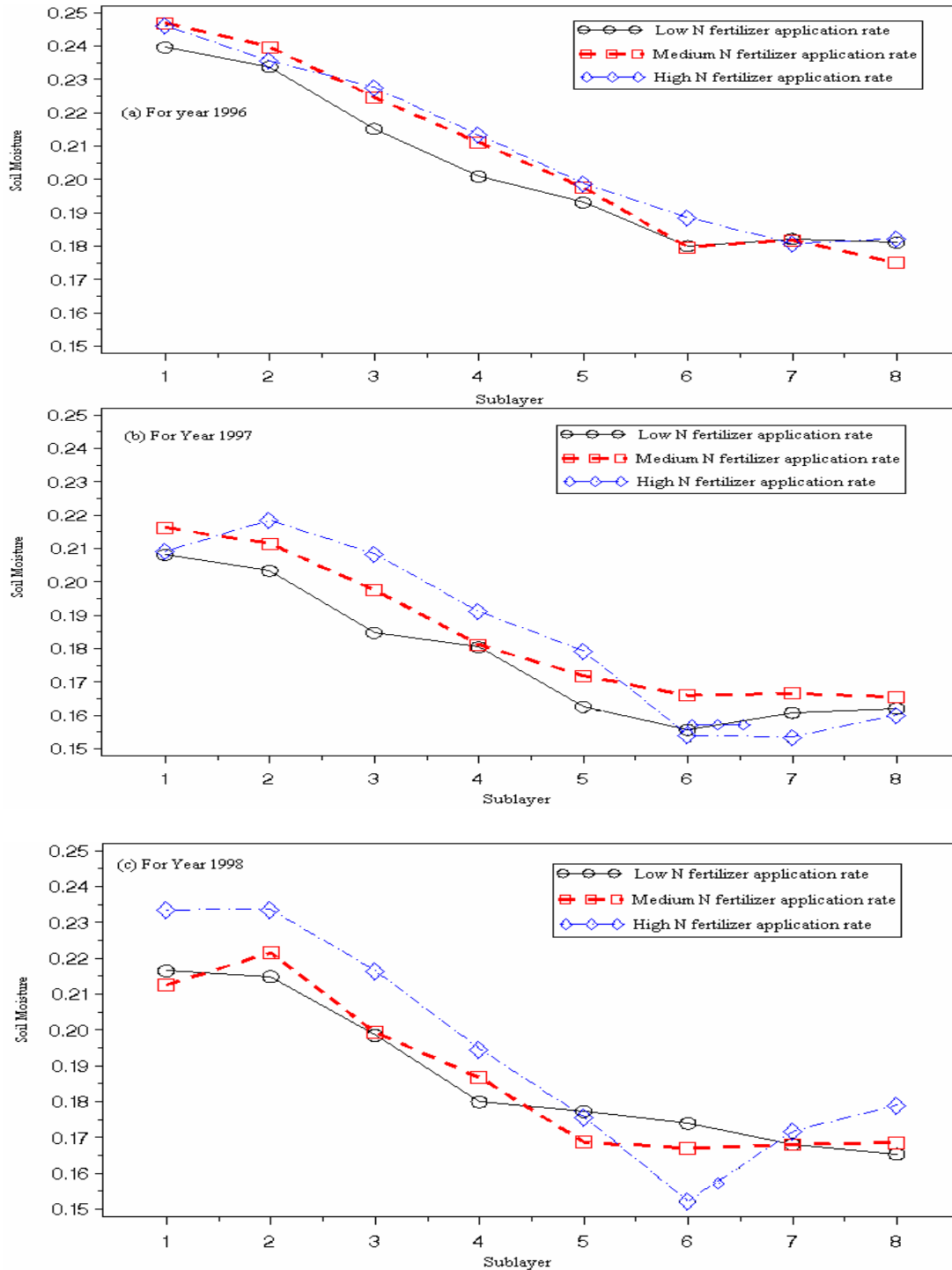


Figure 2. Sample mean plots of soil moisture for the three years and N fertilizer application rates.



Table 1. Test criteria and F approximations for MANOVA Model (1).

Effects & Interactions	Wilks' Lambda	F	df <sub>1</sub> (numerator freedom)	df <sub>2</sub> (denominator degree freedom)	P <sub>r</sub> >F
Apprate <sup>a</sup>	0.9867	2.9736	4	1776	0.0184*
Year <sup>b</sup>	0.8324	42.6460	4	1776	0.0001*
Depth <sup>c</sup>	0.5173	49.5152	14	1776	0.0001*
Apprate/Year	0.9898	1.1438	8	1776	0.3304
Apprate/Depth	0.9834	0.5334	28	1776	0.9786
Year/Depth	0.9687	1.0164	28	1776	0.4413
Apprate/Year/Depth	0.9758	0.3909	56	1776	1.0000

<sup>a</sup>Apprate: N fertilizer application rate.

<sup>b</sup>Year: study year.

<sup>c</sup>Depth: soil depth

\*: significant effect at a significance level of 0.05.

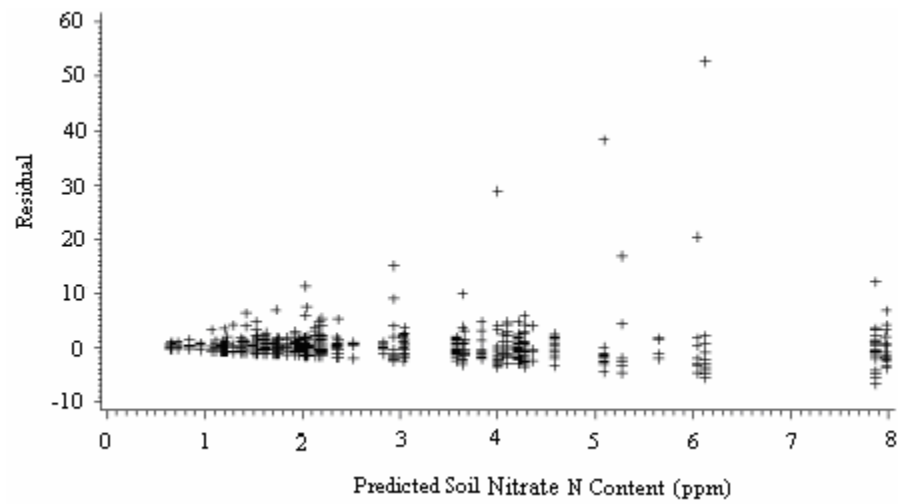
### The Multivariate Linear Regression Model

Both the sample mean plots and MANOVA Model (1) indicate that the main effects of the three factors are significant but the interaction effects insignificant at a significance level of 0.05, i.e., the effects of these factors are additive. Hence, a multivariate linear regression model with independent variables representing these factors can be fit for the dataset. An initial model was developed to include four types of independent variables that represent: 1) N fertilizer application rate; 2) climatic condition; 3) soil depth or sublayer; and 4) crop rotation. However, because the variable representing crop rotation was insignificant at a significance level of 0.15 for model selection (Neter et al., 1996), it was dropped from the model.

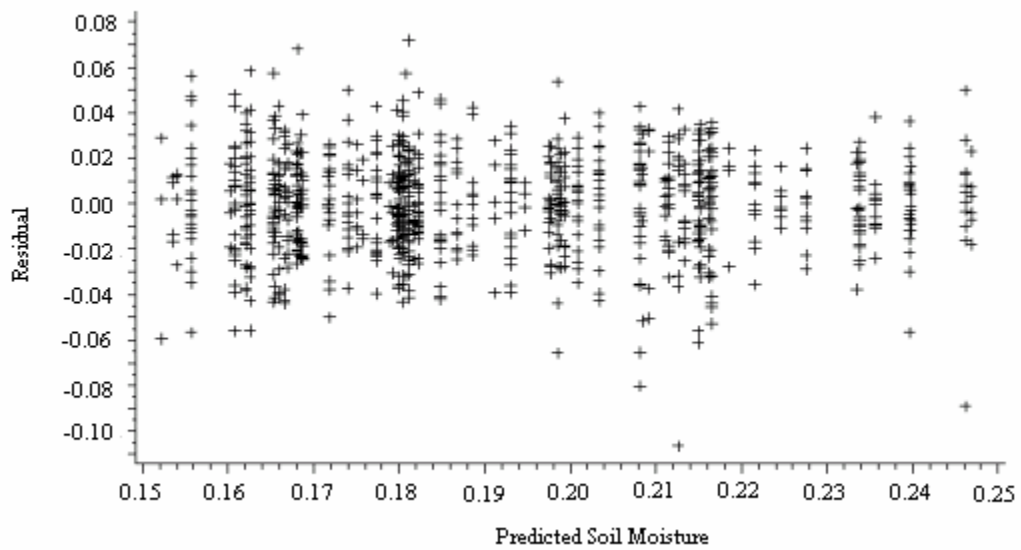
Employing dummy variables, the regression model that does not include the variable for crop rotation can be written as:

$$Y_{962 \times 2} = X_{962 \times 2} \beta_{12 \times 2} + \varepsilon_{962 \times 2} \quad (2)$$

where



(a)



(b)

Figure 3. Residual plots of MANOVA Model (1) for predicting (a) soil nitrate N content and (b) soil moisture.

$$Y_{962 \times 2} = \begin{bmatrix} y_{1,1} & y_{1,2} \\ y_{2,1} & y_{2,2} \\ \dots & \dots \\ y_{962,1} & y_{962,2} \end{bmatrix} \quad \text{— measurement matrix with soil nitrate N content in the first column and soil moisture in the second column}$$

$$\beta_{12 \times 2} = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \\ \dots & \dots \\ \beta_{12,1} & \beta_{12,2} \end{bmatrix} \quad \text{— regression coefficient matrix;}$$

$$\epsilon_{962 \times 2} = \begin{bmatrix} \epsilon_{1,1} & \epsilon_{1,2} \\ \epsilon_{2,1} & \epsilon_{2,2} \\ \dots & \dots \\ \epsilon_{962,1} & \epsilon_{962,2} \end{bmatrix} \quad \text{— random error matrix;}$$

$$X_{962 \times 12} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,11} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,11} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{962,1} & x_{962,2} & \dots & x_{962,11} \end{bmatrix} \quad \text{— design matrix;}$$

$$x_{i,1} = \begin{cases} 1 & \text{low application rate} \\ -1 & \text{high application rate} \\ 0 & \text{medium application rate} \end{cases}$$

$$x_{i,2} = \begin{cases} 1 & \text{medium application rate} \\ -1 & \text{high application rate} \\ 0 & \text{low application rate} \end{cases}$$

$$x_{i,3} = \begin{cases} 1 & \text{dry year (1996)} \\ -1 & \text{normal year (1998)} \\ 0 & \text{wet year (1997)} \end{cases}$$

$$x_{i,4} = \begin{cases} 1 & \text{wet year (1997)} \\ -1 & \text{normal year (1998)} \\ 0 & \text{dry year (1996)} \end{cases}$$

$$x_{i,5} = \begin{cases} 1 & \text{Sublayer 1} \\ -1 & \text{Sublayer 8} \\ 0 & \text{others} \end{cases}$$

$$x_{i,6} = \begin{cases} 1 & \text{Sublayer 2} \\ -1 & \text{Sublayer 8} \\ 0 & \text{others} \end{cases}$$

$$x_{i,7} = \begin{cases} 1 & \text{Sublayer 3} \\ -1 & \text{Sublayer 8} \\ 0 & \text{others} \end{cases}$$

$$x_{i,8} = \begin{cases} 1 & \text{Sublayer 4} \\ -1 & \text{Sublayer 8} \\ 0 & \text{others} \end{cases}$$

$$x_{i,9} = \begin{cases} 1 & \text{Sublayer 5} \\ -1 & \text{Sublayer 8} \\ 0 & \text{others} \end{cases}$$

$$x_{i,10} = \begin{cases} 1 & \text{Sublayer 6} \\ -1 & \text{Sublayer 8} \\ 0 & \text{others} \end{cases}$$

$$x_{i,11} = \begin{cases} 1 & \text{Sublayer 7} \\ -1 & \text{Sublayer 8} \\ 0 & \text{others} \end{cases}$$

$i = 1, 2, \dots, 962.$

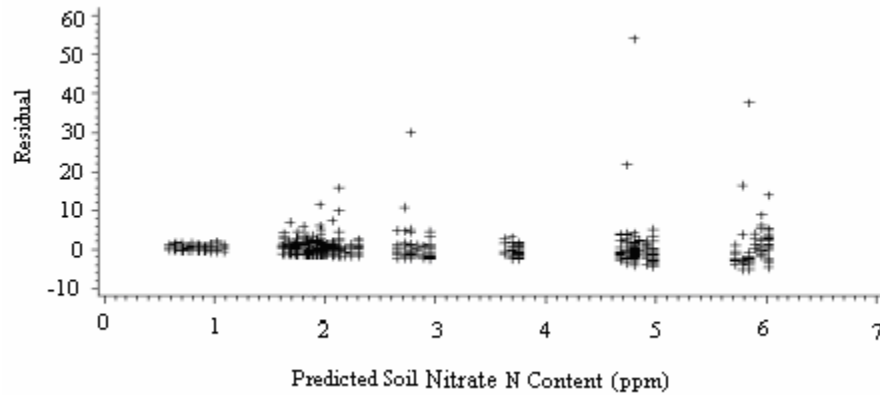
The GLM procedure in the SAS statistical analysis software (SAS Institute Inc., 1996) was used to compute the regression coefficient matrix of Model (2) and MANOVA test statistics, namely Wilks' Lambda, F value, and p value. The MANOVA test indicated that  $x_{i,2}$  was statistically insignificant at any meaningful significance level (p value = 0.95). Dropping  $x_{i,2}$ , the regression coefficient matrix for the final regression model was computed as:

$$\beta = \begin{bmatrix} 2.5292 & 0.1926 \\ 0.0612 & -0.0035 \\ - & - \\ 0.2900 & 0.0140 \\ -0.7548 & -0.0105 \\ 2.9650 & 0.0324 \\ 1.9204 & 0.0294 \\ -0.09859 & 0.0131 \\ -0.7498 & -0.0003 \\ -0.9241 & -0.0121 \\ -1.0743 & -0.0211 \\ -1.1238 & -0.0206 \end{bmatrix}$$

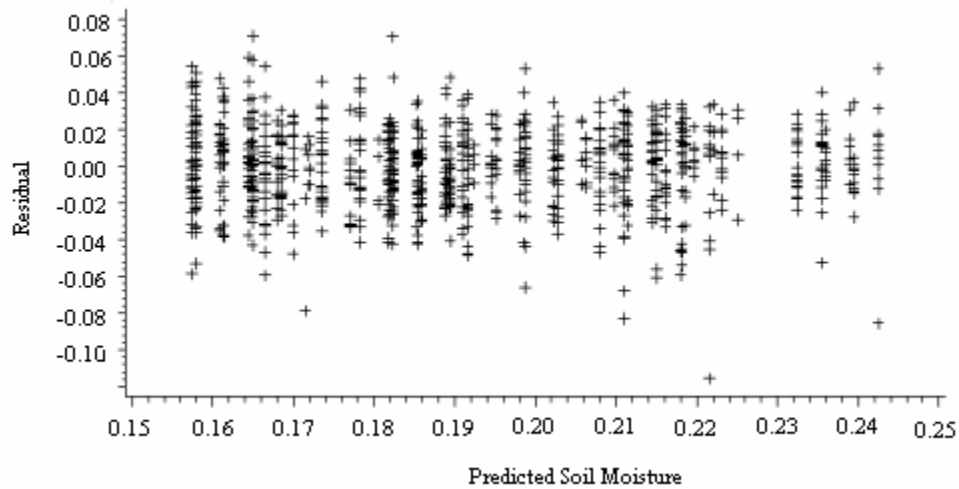
The MANOVA test statistics for this coefficient matrix are listed in Table 2. The p values for all of these coefficients are less than 0.02, indicating that all of them are statistically significant at a significance level of 0.15 for model selection (Neter et al., 1996). As with MANOVA Model (1), the residual plots (Figure 4) indicate that the bivariate normality assumption of Model (2) holds for the dataset.

Table 2. Test criteria and F approximations for the regression coefficient matrix.

Coefficient	Wilks'		df <sub>1</sub> (numerator freedom)	df <sub>2</sub> (denominator freedom)	P <sub>r</sub> >F
	Lamb da	F			
$[\beta_{1,1}, \beta_{1,2}]$	0.015 8	29549. 47	2	949	0.0001
$[\beta_{2,1}, \beta_{2,2}]$	0.985 8	6.8305	2	949	0.0011
$[\beta_{3,1}, \beta_{3,2}]$	—	—	—	—	—
$[\beta_{4,1}, \beta_{4,2}]$	0.836 1	93.012 9	2	949	0.0001
$[\beta_{5,1}, \beta_{5,2}]$	0.872 6	69.285 6	2	949	0.0001
$[\beta_{6,1}, \beta_{6,2}]$	0.699 7	203.66 17	2	949	0.0001
$[\beta_{7,1}, \beta_{7,2}]$	0.767 1	144.03 18	2	949	0.0001
$[\beta_{8,1}, \beta_{8,2}]$	0.951 4	24.242 0	2	949	0.0001
$[\beta_{9,1}, \beta_{9,2}]$	0.991 8	3.9132	2	949	0.0203
$[\beta_{10,1}, \beta_{10,2}]$	0.948 0	26.013 6	2	949	0.0001
$[\beta_{11,1}, \beta_{11,2}]$	0.872 7	69.024 1	2	949	0.0001
$[\beta_{12,1}, \beta_{12,2}]$	0.876 2	67.025 8	2	949	0.0001



(a)



(b)

Figure 4. Residual plots of Model (2) for predicting (a) soil nitrate N content and (b) soil moisture.

**Validation of the Multivariate Linear Regression Model**

As mentioned above, the data measured in 1999 were used to validate Model (2). Figure 5 (a) shows the predicted versus measured soil nitrate N content, and Figure 5 (b) the predicted versus measured soil moisture. The model assumes that a sublayer is homogeneous and does not consider the within-sublayer variations of soil nitrate N content and soil moisture. As a result, the model predicted values represent the average conditions for the sublayer. On the other hand, the measured data include seven to eight samples for each sublayer, which indicate the within-sublayer variations. This is the reason why Figure 5 seems to show same predicted values for different measured soil nitrate N content and soil moisture, as the predicted average conditions were paired with the corresponding seven to eight sampled values. Nevertheless, the coefficients of determination ( $R^2$ ) for predicting soil

nitrate N content and soil moisture were computed to be 0.82 and 0.45, respectively. The high values of  $R^2$  indicate that the model is robust.

To further assess the model prediction performance, for a sublayer, the seven to eight sampled values were arithmetically averaged to compute the corresponding sublayer-averaged value, which in turn was compared with the corresponding predicted value (Figure 6). Overall, the model did a very good job on explaining the variations exhibited by the sublayer-averaged data on soil soil nitrate N content and soil moisture ( $R^2 \geq 0.93$ ). However, the model tended to underestimate the soil nitrate N contents for the upper five sublayers while accurately predicted the ones for the lower two sublayers (Figure 6a). This might be because the soil nitrate N contents of the upper sublayers had a larger within-sublayer variation than that of the lower sublayers, as indicated by the sampled data. This large variation might make average a poor indicator of soil nitrate N content for the upper sublayers. For the same reason, the model tended to underestimate the soil moistures for most of the sublayers (Figure 6b). The model had a slightly better performance on predicting soil moisture, as indicated by the trend line with a slope of 0.95 that is close to 1. For management purposes, this model can be used to predict soil nitrate N content and soil moisture for a location of interest within the study field at various soil depths.

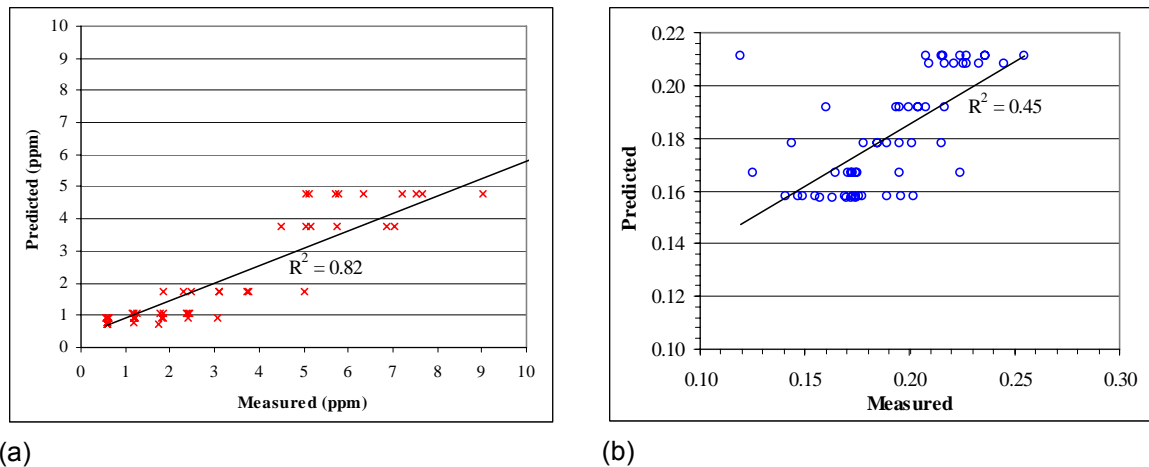


Figure 5. Plots showing predicted versus measured (a) soil nitrate N content and (b) soil moisture for the model validation year of 1999. The measured data include seven to eight samples for a sublayer, whereas, the predicted results are the average values for the sublayer.



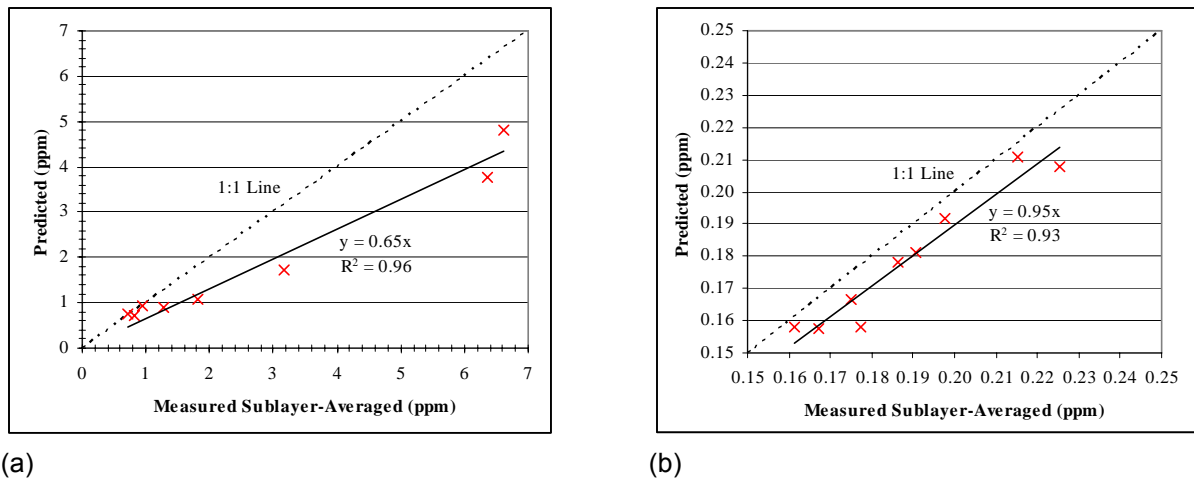


Figure 6. Plots showing predicted versus measured sublayer-averaged (a) soil nitrate N content and (b) soil moisture for the model validation year of 1999. For a sublayer, the measured values for the seven to eight samples were arithmetically averaged to compute the corresponding sublayer-averaged value, which in turn was compared with the predicted value.

### SUMMARY AND CONCLUSIONS

Soil nitrate N is vital to crop growth, but could be a potential nonpoint pollution source when its content is higher than a critical value. Soil nitrate N content is affected by various factors, such as N fertilizer application rate, soil properties, climatic conditions, and cultivation practices (e.g., crop rotation). This study examined how these factors affect soil nitrate N content using multivariate statistical techniques of sample mean plots, MANOVA, and multivariate linear regression.

A dataset of 962 samples, organized as a three-way contingency table, was created using the data measured from 1996 to 1998 in a 25 ha row crop field located in Central Iowa. The sample mean plots generated from this contingency table indicate that the effects of N fertilizer application rate, soil depth, and year on soil nitrate N content are additive. This inference was further verified by a MANOVA model fitted to this dataset. Incorporating the significant effects identified by the sample mean plots and MANOVA model, a multivariate linear regression model was developed. This model considered the variance of, and the correlation between, soil nitrate N content and soil moisture. The validation using the field data collected in 1999 indicated that the model explained more than 93% variations exhibited by the measured sublayer-averaged data on soil nitrate N content and soil moisture. However, this model is unable to account for the within-sublayer variations exhibited by the seven to eight sampled data for a sublayer. Nonetheless, for management purposes, the multivariate regression model can be used to predict soil nitrate N content and soil moisture for the study field as a result of various field management practices and climatic conditions.

## REFERENCES

- Algerbo, P.A., and Thylen, L. (1998). "Variable Nitrogen Application: Effects on Crop Yield and Quality." In Proceedings of the 4<sup>th</sup> International Conference, July 19 – 22, 1998, St. Paul, MN.
- Baker, J.L., and Laflen, J.M. (1982). "Effects of Corn Residue and Fertilizer management on Soluble Nutrient Runoff Losses." *Trans. ASAE* 25: 344 – 348.
- Bakhsh, A., Jaynes, D.B., Colvin, T.S., and Kanwar, R.S. (2000). "Spatio-temporal Analysis of Yield Variability for a Corn-soybean Field in Iowa." *Trans. ASAE* 43(1): 31 – 38.
- CAST (Council for Agricultural Science and Technology). (1992). "Water Quality: Agriculture's Role." Report 120, 23–41. Ames, Iowa: Council for Agriculture Science and Technology.
- Chaney, K. (1990). "Effect of Nitrogen Fertilizer Rate on Soil Nitrate Nitrogen Content After Harvesting Winter Wheat." *J. Agric. Sci.* 114(2): 171 – 176.
- Drury, C.F., Tan, C.C., Gaynor, J.D., Oloya, T.O., and Welacky, T.W. (1996). "Influence of Controlled Drainage–subirrigation on Surface and Tile Drainage Nitrate Loss." *J. Environ. Qual.* 25: 317–324.
- Gentry, L.E., David, M.B., Smith, K.M., and Kovacic, D.A. (1998). "Nitrogen Cycling and Tile Drainage Nitrate Loss in a Corn/Soybean Watershed." *Agriculture, Ecosystems and Environment* 68: 85–97.
- Jaynes, D.B., Colvin, T.S., Karlen, D.L., and Cambardella, C.A. (1999). "Yield and Nitrate Concentration in Subsurface Drainage Water as a Function of N fertilizer Rate." In proceedings of the 1999 ASAE/CSAE–SCGR Annual International Meeting, July 18–21, 1999, Toronto, Ontario, Canada.
- Johnson, R.A., and Wichern, D.W. (1999). *Applied Multivariate Statistical Analysis* (4th ed.). Prentice Hall, Upper Saddle River, New Jersey.
- Kladivko, E.J., and Keeney, D.R. (1987). "Soil Nitrogen Mineralization as Affected by Water and Temperature Interactions." *J. Biology and Fertility of Soils* 5(3): 248 – 252.
- Lykins, B.W., and Clark, R.M. (1995). "U.S. Drinking–water Regulations: Treatment Technologies and Cost." *Journal of Environmental Engineering* 120(4): 783 – 802.
- Marshall, M.G., and Bennett, C.F. (1998). "Outcomes of Nitrogen Fertilizer management Programs." National Extension Targeted Water Quality Program, 1992 – 1995, 3.
- Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. (1996). *Applied Linear Statistical Models* (4<sup>th</sup> ed.). The McGraw-Hill Companies, Inc., NY.
- Olness, A., Lopez, D.R., Cordes, J., Sweeney, C., and Voorhees, W.B. (1997). "A Nitrogen Fertilizer Decision Aid: Theory." *Agron. Abstr.* p.239.
- Olness, A., Evans, S.D., and Alderfer, R. (1998). "Calculation of Optimal Fertilizer Rates: A Comparison of Three Response Models." *J. Agron. Crop Sci.* 180: 215 – 222.
- Paul, K.I., Polglase, P.J., O'connell, A.M., Carlyle, J.C., Smethlirst, P.J., and Khanna, P.K. (2003). "Defining the Relation Between Soil Water Content and Net Nitrogen Mineralization." *Eur. J. Soil. Sci.* 54(1): 39 – 47.
- Richards, I.R., Wallace, P.A., and Paulson, G.A. (1995). "Effects of Applied Nitrogen on Soil Nitrate–Nitrogen Content After Harvest of Winter Barley." *J. Nutrient Cycling in Agroecosystems* 45(1): 61 – 67.

- Sander, D.H., Walters, D.T., and Frank, K.D. (1994). "Nitrogen Testing for Optimum Management." *Journal of Soil and Water Conservation* 49: 46–52.
- SAS Institute Inc. (1996). *SAS User's Manual*. SAS Institute, Cary, NC.
- Selles, F., and James, D. (1999). "Topdressing Nitrogen for Yield and protein management." In *Proceedings of the Dryland Cropping Systems Conference*, March 4 – 5, 1999, Moose Jaw, Saskatchewan.