IDENTIFYING ASSOCIATIONS BETWEEN SOIL AND PRODUCTION VARIABLES USING LINEAR MULTIPLE REGRESSION MODELS

Hari Dahal PhD[*1] and J. K. Routray PhD[2]

ABSTRACT

In agriculture, soil variables are known to be important factors that determine the level of crop productivity in a given situation. To assess which soil variables are important to crop production, soil samples were tested and the test data were correlated with crop yields. A total of six soil variables- soil reaction, organic matter, total nitrogen, available phosphorus, potassium and soil texture were put into Pearson's correlation with crop yield data. Some of the soil variables were found to be highly correlated. To evaluate the apparent strength of the relationship and to explain the variations on dependent variable (crop yield) multiple regression models were developed. In conclusion, it was found that the most important variables explaining the variations in the yield of paddy were total nitrogen, organic matter and phosphorus.

Keywords: correlation, crop yield, linear multiple regression, soil factors

INTRODUCTION

In broad sense, biological production is the product of crop genotype and physical environments. It is therefore integrated product of energy, water, gas, and related biological regimes, which in turn are functions of time. This relationship can be shown as (Nix, 1969);

$$Y(t) = \int_0^t \Delta g(R_e, R_w, R_n, R_g, R_b)\, \Delta t \quad \text{Where,}$$

$$Y \quad = \quad \text{Biological yield over time t}$$
$$\Delta g = \quad \text{Rate of change of biological yield Y that is a function of energy, water,}$$

nutrient, gas, and biotic regimes which are again functions of time, Δt .

In agro ecosystem, in addition to these regimes one more factor plays an important role, this factor is management regime. If crop genotype and all abiotic regimes are taken as environmental factor, then crop yield can be expressed by;

$Y = f(E, M)$ Where,

E = Environmental factor
M = Management factor

Environmental factor broadly conceived as made up of climate and soil factors so that;

$Y = f\{(C+S)\}M$ Where,

Y = Crop yield,
C = Climate,
S = Soil,
M = Management

Since climatic condition in a small area more or less remains uniform, crop yield virtually depends on soil and management practices;

$Y = f(S, M)$

---

[1] Joint Secretary, Ministry of Agriculture and Cooperatives, Kathmandu
[2] Prof., Rur. and Reg. Dev. Planning, Asian Institute of Technology, Thailand.

In this section an effort is made to examine the relationships between crop yield and soil factors keeping the management factor constant. The productivity index model (PIM) which relates crop production to five soil attributes is numerically expressed as used by Pierce and colleagues, 1983 (cited in Younis and Dragun, 1993; NAP, 1993) in the following form:

$$PIM = \sum_{i=1}^{n} (AW_{ci} , AE_{i} , BD_{i} , S_{pH} , E_{cvi} , WF_{i})$$ Where,

AWc	=	sufficiency of available water holding capacity
AE	=	soil aeration
BD	=	bulk density
S pH	=	soil pH
E cv	=	electrical conductivity
WF	=	weighting factor
n	=	number of horizons in crop root distribution
i	=	increments of soil depth

The productivity index is normalized to take values between 0.0 and 1.0. To apply this equation in field conditions is complicated particularly when sufficient information on how productivity is affected by each of these factors individually is not available. Another limitation seems in its greater focus on soil physics rather than on nutritional factors. Therefore, an attempt is made here to see simply how some of the nutrient attributes are linearly related to crop yields for major field crops in the study area. Although a number of soil variables are found keenly related to determining the fertility status of a soil and thus crop productivity for this purpose only six soil variables are assumed to be functionally related to crop yields;

$$Y = f(N, OM, P_2O_5, K_2O, pH, T_x, N_s)\delta t$$ Where,

N	=	Total Nitrogen in the Soils
OM	=	Organic Matter
P_2O_5	=	Available Phosphorus
K_2O	=	Available Potassium
pH	=	Soil pH (a measure of soil reactions)
Tx	=	Textural Class
Ns	=	Non-soil variable, and all these factors are also functions of time as indicated by ∂ t.

MATERIALS AND METHODS

A total of 43 composite soil samples were randomly selected from Babiyabirta VDC of Morang district. To ensure proper area and soil representation the following procedures were adopted:

1. At least five sub-samples were collected from each field unit to make a composite sample.
2. Soil samples were taken at a vertical depth of plow layer (20-25 cm) by spade and auger.
3. The size of composite sample was reduced by successive quartering to about half kilogram.
4. Crop and fertilizer history along with farmers name and address was labeled to the respective sample bag and dispatched to soil laboratory – Regional Agriculture Research Station, Tarhara, Sunsari district.

Nitrogen and organic matter percentage were estimated as a measure of carbon content through Walkley-Black rapid titration method. The phosphorus content kilogram per hectare was analyzed by Olsen's sodium bicarbonate method; potash content kg per hectare was determined by potash turbid metric method and soil reaction by pH meter (1:1 soil-water

ratio). Sand, silt and clay proportion was estimated by hydrometric method of textural analysis. The soil test data is given in the table 1 annex.

The crop yields and other information were assessed through household survey for which a sample size of 187 was calculated using the following formula;

$$n = \frac{N z^2 pq}{N d^2 + z^2 pq}$$, where   N= total population (households), p= expected rate of event

occurrence, q=1-p, z= 1.96 at 95 % confidence level, d= degree of acceptable error and n= sample size. The households were randomly selected and information was gathered through interview using a well-structured questionnaire that was administered to each of the respondents.

RESULTS AND DISCUSSIONS

MEASURES OF RELATIONSHIPS BETWEEN THE SELECTED VARIABLES

To examine the patterns of the relationships between the variables the Pearson's product moment correlation coefficient (r) was used. Since the sample size were more then 30, randomly selected and were in interval-ratio level of measurement Pearson's r thought to be an appropriate tool to measure the association among the soil variables and between the input variables and crop yields.

Although many variables under study may have linear relationships some variables do not exhibit such linearity. The crop yield and added fertilizer nutrient is an example of non linear relationship in which from a maximum point yield starts decreasing, suggesting of a curvilinear relationship. This aspect of relationship is not discussed in this paper. In this section however, we will see the simple form of correlation between two variables as long as the linearity holds good for the data under study. The Pearson product moment correlation coefficient (r) gives both the strength and direction of linear relationship and is expressed by:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$ Where,

   r = Pearson product-moment correlation coefficient, is a measure of the linear
      relationship between two variables such as x and y.
   n = number of observation

A perfect linear correlation is attained when r = $\pm$ 1, and r = 0 implies that x and y tend to have no linear relationship.

The coefficient of correlation indicates how well one variable X is related to the variation in another variable Y. There are a number of ways the coefficient of correlation can be interpreted. One crude method of interpretation is based on the rating scale in which up to 0.20 is considered negligible, from 0.20 to 0.40 is low, 0.40 to 0.60 means moderate, 0.6 to 0.80 is substantial and from 0.80 to 1.0 is considered high to very high (Best and Kahn, 2003).

The interpretation of the correlation becomes easier when explained by the square of the coefficient of correlation, r^2 that is better known by coefficient of determination. The coefficient of determination measures the proportion of variation in one of the variables as explained by the variation in another variable. For an example, if r = .65, then r^2 = .42 as $r = \sqrt{r^2}$. This means that 42 per cent of the total variations in rice yields (Y1) can be explained by the variation in the total nitrogen content (X4) in the soils (Table 1). The remaining 58 per cent variation in rice yield is unexplained by the variation in the nitrogen

content of the soils that means, 58 per cent of the variations in yield could have been caused by the variables other than nitrogen content of the soils. It is however important to note while interpreting that the coefficient of correlation should not be taken as a cause and effect relationship.

TESTING NULL HYPOTHESIS

To see if some soil variables were relevant in crop production in the study area, a null hypothesis (H0:μ) – *there is no relationship between soil variables and the crop yields was tested*. This hypothesis was tested using both correlation (Table 1) and multiple regression models (section 3.5).

The correlation coefficient (Table 1) shows the linear relationship between crop yields and nine soil variables including a non-soil variable-cropping intensity (X_1). Among the eight explanatory soil variables only three variables (organic matter, nitrogen and clay) were significantly related to crop yields whereas none of the variables had significant relationship with maize yield. The yield of monsoon rice (Y_1) found significantly correlated with total nitrogen (X_4), organic matter (X_3), and clay content (X_8) in the soils. It is needless to mention that these variables were positive and the correlation coefficients for X3, and X_4 were at significant at 0.001 levels of probability and for X_8 at 0.05 levels of significance. This means these variables were important and increase in these variables in the soils may enhance the productivity of rice where as soil pH (X_2), sand (X_7), K_2O content (X_6) and P_2O_5 content (X_5) were not significant and inversely related to yield. A weak positive relationship was also observed between silt (X9) and rice yield but no linear relationship was seen between cropping intensity and yields except for jute in which cropping intensity (X_1) was inversely related.

No significant relationship was seen between soil variables and maize yield however a weak positive relationship could be observed between organic matter (X_3), nitrogen (X_4), and silt (X_9) and maize yield whereas negative relationships were seen with pH (X_2), sand (X_7), and clay (X_8). It is obvious that higher percentage of both clay and sand could impede the yield in maize and inverse relationship with pH means less tolerance of maize with more alkaline soils.

Table 1: Correlation coefficients between crop yields and selected explanatory soil variables (2 tailed)

	Rice (Y_1)	Maize (Y_2)	Wheat (Y_3)	Jute (Y_4)	Mustard (Y_5)
Cropping Intensity % (X_1)	0.003	-0.08	0.09	-0.41**	-0.07
pH (X_2)	-0.16	-0.23	-0.09	0.12	-0.23
Organic Matter % (X_3)	0.63*	0.27	0.32***	0.59*	0.40**
Total Nitrogen % (X_4)	0.65*	0.22	0.32***	0.52***	0.37**
P_2O_5 kg ha-1 (X_5)	-0.09	0.05	0.04	-0.07	-0.14
K_2O kg ha-1 (X_6)	-0.03	0.03	0.14	0.05	-0.1
Sand % (X_7)	-0.17	-0.2	-0.04	0.24	0.07
Clay % (X_8)	0.31**	-0.18	-0.11	-0.02	0.12
Silt % (X_9)	0.1	0.27	0.07	-0.26	-0.11

*<P 0.001 **<P 0.01 ***<P 0.05

There was significant relationship observed at 0.05 levels of significance of organic matter (X_3) and total nitrogen (X_4), with wheat yield (Y_3). Soil potassium (X_6) was also found positively related to wheat yield whereas clay, sand, and pH were inversely related.

The yield of jute was significantly correlated with organic matter (X_3) and total nitrogen (X_4) and less closely related with sand (X_7), and pH (X_2). The yield was also strongly but negatively correlated with cropping intensity (X_1), silt (X_9) was also found negatively correlated with jute yield. Mustard yield was also found significantly correlated with organic matter content (X_3) and total nitrogen (X_4). Soil pH (X_2), P_2O_5 (X_5), K_2O (X_6) and silt (X_9) were inversely correlated with the yield whereas other factors were seen not linearly related to the yield of mustard crop.

By and large, Table 1 shows the coefficient of correlation of organic matter and total nitrogen with the yields of all crops except maize were significant at or below 0.05 levels (p-values) indicating these variables were relevant to the yields of major crops. The multiple regression equations for rice especially the last four equations, soil pH, organic matter, total nitrogen, phosphorus and clay content were found significant in explaining the yield variations thus the null hypothesis that the soil variables were irrelevant to the yields of major crops was rejected.

Table 2: Correlation Matrix among selected Soil Variables

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1.0								
X_2	.08	1.0							
X_3	-.08	-.19	1.0						
X_4	-.04	-.18	.95*	1.0					
X_5	.12	.30***	.06	.02	1.0				
6	.13	.11	.11	.07	.32***	1.0			
X_7	-.13	-.05	-.09	-.05	.17	-.01	1.0		
X_8	-.13	-.23	.21	.19	-.36***-	.11	-.44**	1.0	
X_9	.18	.12	.04	.006	-.08	.04	-.96*	.21	1.0

The x notations are the same as those for the table 1, * $P<0.001$ ** $P<0.0\ 1$*** $P<0.05$ (two tailed)

CORRELATION AMONG THE EXPLANATORY VARIABLES

The correlation matrix generated among the soil variables are shown in Table 2. The matrix reveals a high degree of positive correlation between total nitrogen (X_4) and organic matter (X_3), and a high degree of inverse relationship between sand (X_7), and silt (X_9). A strong negative correlation was also observed between sand (X_7) and clay (X_8). A close positive relationship was seen between X_5 (phosphorus) and X_2 ((pH); X_6 (potassium) and X_5 (phosphorus) whereas inverse relationships were observed between X_8 (clay) and X_5 (phosphorus) and between X_8 (clay) and X_2 (pH). The linear correlations among other variables were much weaker.

 MULTIPLE REGRESSION ANALYSIS

The correlation analysis in the preceding section revealed that some of the soil variables were highly interrelated however it is not known how important these variables were in explaining the variations in crop yields. To evaluate the apparent strength of the relationship and to assess the major factors as the determinants of crop yields, multiple linear regression models were developed employing crop yields against the soil variables. The multiple regression model employed here is the extension of simple regression equation such as given below,

$Y = a + bx$ Where,

Y =   dependent variable
a =   y-intercept, the value of which is determined when x = 0
b =   the slope of the regression line or the amount of change produced in Y by a unit change in x
x =   independent variable

The general form of multiple regression model in which more than one independent variables is included is given by,

$$Y = b_o + b_1 x_1 + b_2 x_2 + ... + b_k x_k + e$$

The dependent variable Y is now regarded as a function of k independent variables x_1, x_2, to x_k. The coefficients b_1 to b_k are referred to as the partial regression coefficients which in fact determine the contribution of the independent variables in the equation and b_o is the y-intercept. The random error term (e) that is a random variable assumed to be normally and independently distributed between a mean of zero and a standard deviation of σ that is added to make the model probabilistic.

To determine the usefulness of the entire regression model as how well a linear model fits a set of data the multiple coefficient of determination (R^2) was used.

$$R^2 = \frac{SSR}{SST}$$ Where,   SSR =   regression sum of square or explained variability

SST =   total sum of square or total variability

To know if the regression as a whole is significant or it was due to chance only, F-test was used.

$$F = \frac{MSR}{MSE} = \frac{R^2/k}{(1-R^2)/n-k-1}$$ Where,   MSR =   mean square due to regression

MSE =   mean square due to error

With F distribution having k numerator degrees of freedom (k = number of independent variables) and n-k-1 denominator degrees of freedom, F-test was used in the hypothesis test to determine whether there was a significant relationship among the dependent variable Y and all the independent variable involved in the equation. The hypothesis to test may be,

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$$ (X variables are not useful for predicting crop yield, Y)

$$H_1 :$$ At least one $\beta_1 \neq 0$ (at least one X variable is useful for predicting crop yield, Y)

While performing hypothesis test if the null hypothesis (H_0) is found correct the regression equation cannot be used for prediction of Y. If the null hypothesis is rejected it is clear that one or more partial regression coefficients have a value different from zero. To know which of the coefficients are nonzero and additional tool of t-test is used. While the F test says whether there is a significant relationship between the independent and dependent variables without specifying which variable is significant, the t-test is useful to do a test of significance on each individual independent variable (Hair et al, 2003; Hoshmand, 1988).

$$t = \frac{b_i - \beta_i}{S_{bi}}$$ Where,   t = a t distribution with n-k-1 degrees of freedom

bi = sample regression coefficient
ßi = hypothesized value of population slope for variable i
Sbi = standard error of the regression coefficient bi

Since the hypothesized value $H_0 : \beta = 0$, we can use the following t formula as the test statistic,

$$t = \frac{b_i}{S_{bi}}$$

The other way of looking at the importance of individual variable in the analysis is through standardized coefficients (ß) or beta. These are the coefficients that are obtained by standardizing the variables before running a regression. Putting all the variables on the same scale means making the comparison easier to see which one variable has more effect on the dependent variable. The larger betas are associated with the larger t ratios and lower p-values.

The multiple regression models employed to see the relationship between nine independent variables (x_1 to x_9) and one dependent variable Y (rice yield) is given below:

Y= b_o + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 + e Where,

Y = Crop Yield (kg ha-1)	X_5= P_2O_5 kg ha-1
b_o= Intercept value	X_6= K_2O kg ha-1
X_1= Cropping Intensity (%)	X_7= Sand (%)
X_2= pH value (in pH scale)	X_8= Clay (%)
X_3= Organic Matter Content (%)	X_9= Silt (%)
X_4= Total Nitrogen Content (%)	e = Error term

$b_1...b_9$= Corresponding coefficients of X1 through X9.

THE RICE MODELS

The regression model yielded the following equation while regressed to the nine explanatory variables against the rice yield;

YR_1 = 487.88 + 0.693X_1 + 6.24X_2 + 144.15X_3 + 1084.15X_4 - 0.163X_5 - 0.138X_6 + 41.68X_7 + 3.25X_8
            (0.51)   (0.05)   (0.32)   (1.16)   (-0.31)   (-0.49)   (1.13)   (0.34)
        R^2 = 0.4788
        F = 3.90 (0.0023)

Although F ratio (figure in parenthesis indicates significance level) was significant indicating the overall explanatory power of this equation and regression line explaining about 47.88 percent (R^2 = 0.4788) variations in rice yield by the independent variables taken together none of the coefficients were significant (t-value in parentheses). The low t-values indicate the impacts of these predictor variables on rice yield were not significantly different from zero. This means the model had no satisfactory predictive ability due to the inclusion of some unimportant or redundant explanatory variables in the equation. To screen such variables in the equation and to improve the model stepwise regression analysis was employed which on iteration yielded the following equation;

YR_2 = 1112.95 + 14242.69X_4
                (5.56)

        R^2 = 0.4298
        F = 30.91 (0.0000)

In this model both F and t values were improved significantly over the previous model, which led to state that total nitrogen (X_4) alone was accounted for about 42.98 % variations in rice yield. The significant positive relation suggests the increase in the amount of nitrogen in the soils also increases the yield of rice, some 57 % variations in the yield left

unexplained by this model that was attributed to other factors that could not be measured or included.

From the Table 2 it was clear that variable X_3 (organic matter) and X_4 (total nitrogen) were highly correlated (r=0.95), similarly a high degree of negative correlation (r=-0.96) was observed between X_9 (silt) and X_7 (sand). The problem of multicollinearity could be encountered when two or more explanatory variables are highly correlated with each other implying the regression model unable to identify the effect of each individual variable on the criterion variable. Although R^2 was far less than r, t value was significant and standard error of the equation YR_2 was relatively small still multicollinearity could not be completely dismissed. More importantly the overall explanatory power (R^2=0.4298) was not to the expectation it was chosen therefore to develop a number of regression models by manipulating (adding1 and dropping) some of the X-variables to understand the importance of the individual coefficient explaining the variation on the criterion variable Y.

$$YR_3 = -474.14 + 41.1X_2 + 1350.52X_4 - 0.148X_5 + 8.475X_8 + 2.075X_{10}$$
$$\quad\quad\quad (4.08) \quad (3.81) \quad\quad (-2.67) \quad (3.86) \quad (15.60)$$

$$R^2 = 0.9866$$
$$F = 74.02 \ (0.0001)$$

When dropping out two variables X_3 (organic matter)and X_7 (sand) which were highly correlated with other two explanatory variables respectively and adding one more variable X_{10} (total land holding) in the original model five variables appeared in the equation (YR_3) explaining about 99% variations in the mean yield of rice. Among the coefficients, land holding (X_{10}) appeared to be most important variable followed by pH value (X_2). The explanation to this may be that the yield was linearly related to the size of land holding (X_{10}) implying larger the land sizes higher the total production (but not necessarily yield per unit of land). The second important variable X_2 (pH) was positively related which means larger the value (more alkaline soil) higher the production which in fact was not relevant to rice crop due to theoretical ground. Therefore, the sign for variable X_2 (pH) was unexpected in the equation. Total nitrogen (X_4), and clay (X_8) were positive implying these variables were relevant to rice crop, higher nitrogen content positively contributes to yield, and clay (X_8) had positive impact on rice production by improving the water holding capacity of the soils. Since the P_2O_5 content in the soils in most cases was high, the negative sign for the variable X_5 was not irrelevant in the equation. When dropping out the X_1 (cropping intensity) in the equation YR_3 yielded the following model:

$$YR_4 = -89.01 + 1016.1X_4 + 1.90X_{10}$$
$$\quad\quad\quad\quad (3.39) \quad (14.27)$$

$$R^2 = 0.8442$$
$$F = 108.44 \ (0.0000)$$

This equation again shows the importance of variable X_{10} (total land holding)and X_4 (total nitrogen) both of which together explaining about 84 % (slightly lower than in YR3) variation in the yield. Some of the regression models developed through the various combinations of explanatory variables are:

$$YR_5 = -68.3 - 0.515X_1 + 916.96X_4 + 3.35X_9 + 0.26X_{11} + 2.24X_{13}$$
$$\quad\quad\quad\quad (-2.06) \quad (1.99) \quad\quad (2.29) \quad (5.11) \quad (4.01)$$

$$R^2 = 0.7347$$
$$F = 16.06 \ (0.0000)$$

5 Added variables were X_{10} (total land holding), X_{11} (total fertilizer used), X_{12} (total FYM), and X_{13} (area under legume crops). The possible impact of variables X_{11} and X_{12} were only to see their residual effects on soil fertility because these were applied in the soils to different crops for the previous year.

$YR_6 = -5297.14 + 9.41X_1 + 41641.95X_4 + 75.73X_{10}$
              (1.69)   (3.55)   (14.45)

       $R^2 = 0.8549$
       F = 76.62 (0.0000)

$YR_7 = -17473.47 + 1685.23X_2 + 2207.02X_3 - 13.0X_5 + 237.19X_8 + 95.7X_{10}$
                (5.05)   (4.66)   (-5.86)   (2.76)   (19.5)

       $R^2 = 0.9902$
       F = 101.5 (0.0001)

$YR_8 = -87.25 + 1169.28X_4 - 0.318X_5 + 1.88X_{10} + 0.952X_{13}$
                (3.68)   (-3.22)   (10.29)   (2.12)

       $R^2 = 0.8693$
       F = 49.88 (0.0000)

$YR_9 = -89.39 + 59.73X_3 - 0.36X_5 + 1.88X_{10} + 0.952X_{13}$
                (4.08)   (-3.7)   (10.71)   (2.52)

       $R^2 = 0.8779$
       F = 53.93 (0.0000)

$YR_{10} = -87.25 + 1169.28X_4 - 0.318X_5 + 1.86X_{10} + 0.824X_{13}$
                 (3.68)   (-3.22)   (10.28)   (2.12)

       $R^2 = 0.8693$
       F = 49.88 (0.0000)

In equation YR_5 nine variables (excluding X_3 (organic matter) and X_7 (sand) and adding X_{13} (area under legume crops) were regressed giving rise to five variables explaining about 73 % variations in the yield. All variables were positively related to yield except variable X_1 (cropping intensity) that had negative sign meaning higher the cropping intensity lower the productivity (unless extra nutrient inputs are added). This relationship seems to be working for all crops under low input cropping systems. The deficit nutrient nitrogen (X_4), and silt (X_9) were appeared in the equation but they seemed less important than non-soil variables X_{11} (total fertilizer used) and X_{13} (legume area).

In YR_6 two non-soil variables (X_{11}, and X_{13}) included in equation YR_5 were dropped and one non-soil variables X_{10}, and two previously dropped variables X_3 and X_7 were included resulting in three variables X_1, X_4 and X_{10} explaining about 85 % variations in yield. In the equation total nitrogen (X_4) appeared as an important variable next to total land holding (X_{10}).

In YR_7 and YR_8 models it is striking to note that variable X_{10} appeared very important having largest t-value followed by a positive X_2, variable X_5 remained important and negative. All other variables were positive and important, altogether explaining about 99 % and 87 % variations in the yield in both equations YR_7 and YR_8 respectively. YR_9 and YR_{10} have similar patterns having phosphorus (X_5), land holding (X_{10}) and legume crop area (X_{13}) as their important variables.

FINDINGS AND CONCLUSIONS

In summing up, analyzing the equations it could be said that the critical soil variables explaining the most variations in the yield of rice were total nitrogen (X_4) and organic matter (X_3). Phosphorus (X_5), soil pH (X_2) and clay (X_8) were also turned out to be important variables in some equations. All these variables but one phosphorus (X_5) was positive. P_2O_5 (X_5) was inversely related to yield in almost all equations that it appeared. In most equations intercept (b_o) was less meaningful when X = 0.

The soil analysis has revealed that the fertility status of soils in the study area was low when considered the levels of two most critical soil variables – total nitrogen and organic matter content. This information is important particularly when soil test values are calibrated to recommend the optimum level of fertilizers (nutrients) in the local farming system. The distribution of phosphorus content in the soil in general was from low to very high level (annex table 1, very high standard deviation) with a mean of high. The reason behind the inverse relationship of phosphorus with the yields may have been due to having high to very high levels of phosphorus content in the soils which might have affected yields negatively.

In view of the findings and conclusions, it is recommended that the policy and improved technological options must be made available to the farmers to conserve and enhance the organic matter contents and total nitrogen in the soils. Wider information dissemination on the methods of balanced use of chemical fertilizers is also needed to maintain soil fertility and improving crop production in the study area.

ACKNOWLEDGEMENT

REFERENCES

Best, W. B. and J. V. Kahn, 2003. Research in Education. New Delhi: Printice-Hall of India Pvt-Ltd.

Dahal H., 1996. Ecological approach to sustainable agriculture through integrated nutrient resource management: a micro level study in the eastern terai farming system, Nepal. PhD dissertation submitted to School of Environment, Resources and Development, Asian Institute of Technology, Thailand.

Hair J. F., R. E. Anderson, R. L. Tatham and W. C. Black (eds.), 2003. Multivariate Data Analysis (fifth edition). Singapore: Pearson Education.

Hoshmand A. R., 1988. Statistical Methods for Agricultural Sciences. California State Polytechnic University, Pomona. Timber Press, Portland, Oregon, USA.

NAP, 1993. Soil and Water Quality: An Agenda for Agriculture. Washington DC: National Academic Press.

Nix, H. A., 1968. Assessment of biological productivity. **In**: G. A. Stewart (ed.), Land Evaluation (Papers of a CSIRO Symposium). Macmillan of Australia.

Younis A. S. and A. K. Dragun, 1993. Land and Soil Management. Western Press, Oxford and IBH Publishing Company Pvt Ltd.

ANNEX

Table 1: Soil test data

Soil Variables	Mean	Standard Deviation	Index mean Value
pH	6.201	0.601	6.41 (slightly acidic)
Organic Matter (%)	2.036	0.510	2.21 (low)
Total Nitrogen (%)	0.102	0.024	0.108 (low)
Phosphorus (kg/ha)	123.97	147.18	77 (high)
Potassium (kg/ha)	296.74	255.6	254.5 (high)
Sand (%)	56.07	7.99	
Silt (%)	32.63	7.32	
Clay (%)	11.30	1.96	