바이오시스템공학 (*J. of Biosystems Eng.*)
Vol. 30, No. 1, pp. 45~55 (2005)

벼와 옥수수 재배 포장에서 경로분석을 이용한 작물 수확량 제한요인 분석

정선옥　K. A. Sudduth　장영창

Path Analysis of Factors Limiting Crop Yield in Rice Paddy and Upland Corn Fields

S. O. Chung　K. A. Sudduth　Y. C. Chang

Abstract

Knowledge of the relationship between crop yield and yield-limiting factors is essential for precision farming. However, developing this knowledge is not easy because these yield-limiting factors are interrelated and affect crop yield in different ways. In this study, data for grain yield and yield-limiting factors, including crop chlorophyll content, soil chemical properties, and topography were collected for a small (0.3 ha) rice paddy field in Korea and a large (36 ha) upland corn field in the USA, and relationships were investigated with path analysis. Using this approach, the effects of limiting factors on crop yield could be separated into direct effects and indirect effects acting through other factors. Path analysis provided more insight into these complex relationships than did simple correlation or multiple linear regression analysis.

Results of correlation analysis for the rice paddy field showed that EC, Ca, and SiO_2 had significant ($P<0.1$) correlations with rice yield, while pH, Ca, Mg, Na, SiO_2, and P_2O_5 had significant correlations with the SPAD chlorophyll reading. Path analysis provided additional information about the importance and contribution paths of soil variables to rice yield and growth. Ca had the highest direct effect (0.52) and indirect effect via Mg (-0.37) on rice yield. The indirect effect of Mg through Ca (0.51) was higher than the direct effect (-0.38). Path analysis also enabled more appropriate selection of important factors limiting crop yield by considering cause-and-effect relationships among predictor and response variables. For example, although pH showed a positive correlation ($r=0.35$) with SPAD readings, the correlation was mainly due to the indirect positive effects acting through Mg and SiO_2, while pH not only showed negative direct effects, but also negatively impacted indirect effects of other variables on SPAD readings.

For the large upland Missouri corn field, two topographic factors, elevation and slope, had significant ($P<0.1$) direct effects on yield and highly significant ($P<0.01$) correlations with other limiting factors. Based on the correlation analysis alone, P and K were determined to be nutrients that would increase corn yield for this field. With the help of path analysis, however, increases in Mg could also be expected to increase corn yield in this case. In general, path analysis results were consistent with published optimum ranges of nutrients for rice and corn production. We conclude that path analysis can be a useful tool to investigate interrelationships between crop yield and yield limiting factors on a site-specific basis.

Keywords : Precision agriculture, Limiting factor, Crop yield, Path analysis

1. INTRODUCTION

The underlying concept of precision agriculture is to recognize the existence of spatial variability in site variables such as yield, soil properties, environmental factors, and topography within fields, and to manage this variability

The article was submitted for publication in December 2004, reviewed and approved for publication by editorial board of KSAM in January 2005.
The authors are Sun-Ok Chung, KSAM member, Agricultural Engineering Researcher, National Institute of Agricultural Engineering, Suwon, Republic of Korea, Kenneth A. Sudduth, Agricultural Engineer, USDA-ARS Cropping Systems and Water Quality Research Unit, University of Missouri, Columbia, Missouri, USA, Young-Chang Chang, KSAM member, Teaching Professor, Konkuk University, Chungju, Korea. The corresponding author is Sun-Ok Chung, National Institute of Agricultural Engineering, Rural Development Administration, 249 Seodun-dong Kwonsun-gu, Suwon, 441-707, Republic of Korea; E-mail : <sochung@rda.go.kr>

according to site-specific conditions and requirements in order to optimize crop yield and minimize environmental damage. The cycle of precision agriculture, or site-specific crop management (SSCM), requires: (1) intensive data collection, (2) decision making or management planning, (3) precision field operations and (4) evaluation (Sudduth, 1998). Understanding the response of crop yield to yield-limiting factors and the interrelationships among the factors is critically important for successful SSCM (Sudduth et al., 1996).

Relationships between crop yield and yield-limiting factors have traditionally been investigated by means of numerous small plot trials over multiple years. However, this approach assumes that all yield-limiting factors are included in the process and the response curves are generally based on test results averaged over areas such as a soil series, a single field, or test plots (Kitchen et al., 1999). With precision agriculture technologies, it is possible to treat each individual geo-referenced data location or small area within a field as a "test plot". The availability of this relatively large amount of data in SSCM has stimulated various approaches for understanding the crop response to yield-limiting factors. These approaches generally have involved either the application of crop growth models (Fraisse et al., 2001a; Irmak et al., 2001) or multivariate statistical analysis (Drummond et al., 2003; Fraisse et al., 2001b; Sudduth et al., 1996).

Scatter plots and correlation analysis can be used to visualize overall relationships among field variables, while the multiple linear regression approach provides a relationship between a response variable (i.e., crop yield) and predictor variables (i.e., yield-limiting factors) (Cambardella et al., 1996; Mallarino et al., 1996). These methods are relatively easy to apply, but they are based on assumptions that: (1) relationships between variables can be parameterized in a suitable form and (2) yield-limiting factors are independent of one another. These assumptions are generally not true for spatial field data, leading to relatively small correlation coefficients and coefficients of determination.

Because the relationships between crop yield and yield-limiting factors can be nonlinear and complex, nonlinear models and parametric nonlinear regression may provide better results than linear methods. The most commonly accepted models for nutrient response are linear-plateau, exponential and quadratic models (Tisdale et al., 1999). These nutrient response models, however, only represent the response of crop yield to the quantitative change of a single factor.

More flexible approaches such as the boundary line method, projection pursuit regression and neural network analysis have been demonstrated. The boundary line method extracts a response curve at the upper edge of any body of data, assuming that the curve defines the best performance in the population. Data below the boundary line represent locations where other, unmodeled factors are limiting yield (Kitchen et al., 1999; Webb, 1972). Both projection pursuit regression and neural network analysis provide general, nonlinear, nonparametric approaches to data analysis. Projection pursuit regression represents a dependent variable as the sum of a set of general (nonlinear) functions of linear combinations of the independent variables, while neural network analysis uses simple, highly interconnected multiple layers that produce an output by responding to a set of inputs (Drummond et al., 2003; Sudduth et al., 1996).

One approach to address the problem of interrelationships among yield-limiting factors is principal component analysis (e.g., Dobermann, 1994; Fraisse et al, 2001b). The principal component approach transforms original, correlated data sets into new uncorrelated data sets, called principal components, which are linear combination of the original variables (Jensen, 1996). Investigating the loading of the original variables to the principal components can help to organize the original independent variables into groups that behave similarly. However, after the principal component transformation is applied it is often not easy to perceive the response of yield to the original variables.

As another approach to deal with correlated data, path analysis (e.g., Gravois and Helms, 1992) provides insights into correlation structures among variables using both correlation and multiple linear regression analyses. Path analysis differentiates between direct effects of an individual predictor variable and indirect effects of a predictor variable acting through other variables on a response variable (Williams et al., 1990). The use of path analysis is expanding in the biological and agricultural sciences because of the insights it can provide into correlation structures among variables (Kasap et al., 1999; Krishnasamy and Mathan, 2001; Pantone et al., 1992; Ssango et al., 2004). Examples of previous path analysis applications include:

1. Effects of forage constituents, quantity of available forage, and forage nutrients (N, P, K, and S) on the per-acre weight gain of lambs (Williams et al., 1990),
2. Evaluation of the effect of soil pH, cation exchange capacity (CEC), organic carbon content and clay content on adsorption of Cd, Cu, Ni, Pb, and Zn by soils (Kasap et al., 1999),
3. Interrelationships among rice yield and yield components for direct-seeded rice cultural systems (Gravois and Helms, 1992), and
4. Evaluation of the competitive interaction between a weed (red rice) and cultivated rice (Pantone et al., 1992).

The overall objective of this study was to apply path analysis to two sets of spatial data, including crop yield and soil and plant properties. Specific objectives were: (1) to evaluate spatial yield response to site variables, and the interrelationships among those variables in terms of path analysis "direct effects" and "indirect effects", and (2) to discuss the use of path analysis for precision agriculture and its potential to analyze nonlinear interrelated factors.

2. MATERIALS AND METHODS

A. Data Acquisition and Processing

Two geo-referenced data sets composed of crop yield and soil and plant properties were collected and processed. Data were obtained from two fields: one located near Suwon in the mid-west part of the Republic of Korea (field 1; 37.284 N, 126.956 E; 0.3 ha, 100 m by 30 m) and the other located near Centralia, in central Missouri, USA (field 2; 39.230 N, 92.117 W; 35 ha, 790 m by 455 m).

Data for field 1, a rectangular, rice (*Oriza Sativa* L.) paddy field, were collected in 1999. Soil classification for this field was Coarse loamy, mixed nonacid, mesic family of Aguic Fluventic Eutrochrepts (Institute of Agricultural Science, 1984). Rice was transplanted in late May, and fertilizers were applied in early April and late July according to RDA recommendations (NIAST, 1999). Data collected from field 1 were rice yield, crop growth status, and basic soil properties. Rice yield (ton/ha) was manually collected on October 13, 1999. The sampling grid spacing was 10 m by 5 m. Yields were determined on 5 stalks, collected at each of three locations in each grid cell. A total of 180 yield samples (60 grid cells x 3 locations) were collected. The samples were threshed with a wooden rice thresher, and the weight and moisture content of each sample were measured with an electronic scale and a moisture meter. Chlorophyll content, an indication of the growth status of the crop, was measured on June 6, 1999, before heading of the rice. Data were collected with a SPAD 502 meter[1)] (Minolta Camera Corp., Japan) on a 2-m by 2-m grid. In each of the 750 cells, 30 SPAD readings were obtained and averaged.

Althoughore planting would be a more standard timing, in this study soil samples were collected to a depth of 15 cm with a spade on October 28, 1999. Thus, the samples reflected the effects of fertilizer additions throughout the season and nutrient removal by the crop, and provided an indication of whether fertility levels dropped below the optimum range during the growing season. Soil sampling was done on three different scales: (1) a 10-m by 5-m grid covering the entire field (60 samples); (2) an intensive 1-m by 1-m grid imposed at two 10-m by 10-m locations at the edge and center of the field (200 samples); and (3) a coarse 20-m by 10-m grid covering the entire field (15 samples). Laboratory analysis was completed by the Soil Management Division, National Institute of Agricultural Science and Technology, Rural Development Administration, for the following properties: pH, electrical conductivity (EC, dS/m), organic matter content (OM, %), P_2O_5 (ppm), Ca (cmol/kg), K (cmol/kg), Mg (cmol/kg), Na (cmol/kg), total nitrogen (N, %), and SiO_2 (ppm). More details on the data collection in field 1 are available in Chung et al. (2000) and Sung et al. (1999).

The soils found at field 2 were of the Mexico series (fine, smectitic, mesic aeric Vertic Epiaqualfs) and the Adco series (fine, smectitic, mesic aeric Vertic Albaqualfs). These soils were formed in moderately-fine textured loess over a

1) Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by NIAE, Korea or USDA-ARS, USA.

fine textured pedisediment and were classified as somewhat poorly drained. Surface textural classes ranged from silt loam to silty clay loam. The subsoil claypan horizon (s) were silty clay loam, silty clay or clay, and commonly contained as much as 50 to 60% smectitic clay. Topsoil depth above the claypan (depth to the first Bt horizon) ranged from less than 10 cm to greater than 100 cm. The field had been managed in a minimum-tillage corn-soybean rotation since 1990 (Sudduth et al., 2003).

Corn yield data for field 2 were obtained in 1997 using a full-size combine equipped with a commercial yield monitoring system and global positioning system (GPS) receiver, using data collection and processing techniques described by Birrell et al. (1996). Elevation and slope data were obtained using a Nikon Topgun A200LG total station surveying instrument (accuracy < 1 cm) and standard mapping procedures. For soil properties, composite soil samples to a 20-cm depth were collected on a 30-m grid in the spring of 1995. Three soil cores obtained within a 1-m radius of each sample position were combined, oven dried, and analyzed by the University of Missouri Soil and Plant Testing Services Laboratory. Soil properties measured were P (ppm), K (ppm) , pH, OM (%), Ca (ppm), Mg (ppm), and cation exchange capacity (CEC, meq/100 g). Apparent electrical conductivity of the soil (EC_a) was measured in November of 1997 using a Geonics EM38 (Geonics Limited, Mississauga, Ontario, Canada). The EM38 was operated in the vertical dipole mode, using a mobile system and data collection procedures described by Sudduth et al. (2001), providing an effective measurement depth of approximately 1.5 m. More details on data collection in field 2 are available in Drummond et al. (2003).

Not all variables were collected on the same spatial grid. Therefore, the nearest neighbor method was used to merge all observations with the most sparsely collected data. A total of 87 and 336 observations were obtained for fields 1 and 2, respectively. Means and standard deviations for the data sets, and optimum ranges of soil parameters for paddy rice (NIAST, 1999) and corn (Buchholz, 1983) are summarized in Tables 1 and 2. For rice, variables with mean values below the optimum range were pH, OM, Ca, Mg, and SiO_2, and variables with mean values above the optimum range were P_2O_5, K, N, and EC. For corn, variables with mean values below the optimum range were pH, P_2O_5, Mg, and K.

B. Analytical Procedures

Path analysis is a statistical technique that differentiates between correlation and causation (Kasap et al., 1999) using both multiple linear regression and correlation analysis. An important condition for path analysis is that a cause-and-effect relationship between predictor variables and the response variable is implicit. Researchers should have the causal hypothesis, between crop yield and limiting factors in this case, to run a path analysis.

Path analysis is based on interpretation of the normal equations used to solve for standardized partial regression coefficients in multiple regression problems. Information obtained from a correlation analysis can be augmented by partitioning the overall effects (described by the correlation coefficient) of predictor variables on a response variable into direct effects and indirect effects for a given set of cause-and-effect interrelationships (Gravois and Helms, 1992). The direct effect is the partial regression coefficient of a predictor variable (e.g., Mg in the soil) for a response variable (e.g., rice yield). Indirect effects are those effects of a predictor variable (e.g., Mg) on a response variable (e.g., rice yield) acting through other predictor variables (e.g., pH), and consist of correlation coefficients with the other predictor variables (e.g., Mg vs. pH) multiplied by the corresponding partial regression coefficients for the response variable (e.g., Mg vs. rice yield). A summation of the direct effect and indirect effects results in the correlation coefficient. For example, the overall effect of Mg on rice yield (described by the correlation coefficient) can be separated into a direct effect (partial regression coefficient) and indirect effects through other predictor variables such as pH.

Williams et al. (1990) provided an example of path analysis for five predictor variables (x_1-x_5) and one response variable (x_6), and described the steps in path analysis:

1. Drawing a path diagram (Fig. 1) from cause-and-effect relationships between predictor variables (x_1-x_5, soil properties and topographic variables in this case) and a response variable (x_6, crop yield in this case). Single-headed and double-headed arrows indicate direct and indirect effects, respectively.

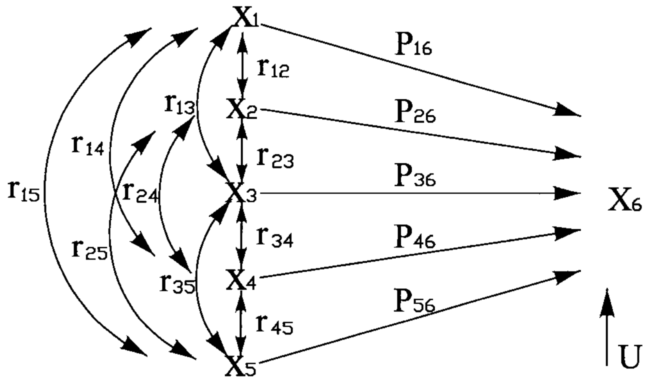

Fig. 1 Path diagram for five predictor variables and one response variable. The variable U is the undetermined portion or residual, $\sqrt{1-R^2}$ (after Williams et al., 1990).

2. Standardization of the variables, centered with means equal to zero and scaled in standard deviation units, since path analysis is based on the correlation structure.
3. Computation of standard partial regression coefficients, called path coefficients (P_{ij}) indicating direct effects, and correlation coefficients (r_{ij}).
4. Completion of the path table in the form of normal equations (equations 1 to 5). Use partial regression coefficients for direct effects, and product of the direct effect and correction coefficient for the indirect effect of each predictor variable on the response variable.

$$p_{16}+r_{12}p_{26}+r_{13}p_{36}+r_{14}p_{46}+r_{15}p_{56}=r_{16} \tag{1}$$

$$r_{21}p_{16}+p_{26}+r_{23}p_{36}+r_{24}p_{46}+r_{25}p_{56}=r_{26} \tag{2}$$

$$r_{31}p_{16}+r_{32}p_{26}+p_{36}+r_{34}p_{46}+r_{35}p_{56}=r_{36} \tag{3}$$

$$r_{41}p_{16}+r_{42}p_{26}+r_{43}p_{36}+p_{46}+r_{45}p_{56}=r_{46} \tag{4}$$

$$r_{51}p_{16}+r_{52}p_{26}+r_{53}p_{36}+r_{54}p_{46}+p_{56}=r_{56} \tag{5}$$

In developing the path analysis for this study, crop yield was the response variable and limiting factors such as soil properties and topographic attributes were used as predictor variables. Path analysis with the SPAD 502 reading as a response variable was also conducted for field 1. The path analysis method followed the above outline. First, all the variables were standardized with means equal to zero and standard deviations equal to one. Then, multiple regression was applied to calculate the partial regression coefficients for predictor variables. Pearson correlation coefficients were obtained among the variables. Finally, path tables were computed in the form of normal equations. Standardized data, Pearson correlation coefficients, and standardized partial regression coefficients were obtained using STAND, CORR, and REG procedures in SAS version 8.2 (SAS Institute Inc., Cary, N.C.). Instead of using all the predictor variables, the forward variable selection option was used to reduce multicollinearity among the variables and to enhance reliability of the models, and the default significance level for entry into the model (SLENTRY=0.5) was used.

Table 1 Means and standard deviations of measured variables for field 1 (n=87), and optimum ranges of soil parameters for rice production.

	pH	OM (%)	P_2O_5 (ppm)	Ca	Mg	K	Na	N (%)	EC (dS/m)	SiO_2 (ppm)	SPAD	Yield (Mg/ha)
				---------(cmol$^+$/kg)---------								
Mean	5.9	2.4	120	4.70	0.98	0.68	0.29	0.2	0.43	93.5	36.6	6.1
Std.	0.2	0.2	14	0.71	0.16	0.10	0.06	0.1	0.12	23.6	1.0	1.0
Optimum range[a]	6.0-7.0	2.5-3.0	80-120	5.0-6.0	1.5-2.0	0.25-0.30			<0.2	<0.2	130-180	

[a] Ranges were obtained from NIAST (1999).

Table 2 Means and standard deviations of measured variables for field 2 (n=336), and optimum ranges of soil parameters for corn production.

	pH	OM (%)	P_2O_5 (ppm)	Ca	Mg	K	CEC (meq/100 g)	EC_a (dS/m)	Elevation (m)	Slope (%)	Yield (Mg/ha)
				---------(cmol$^+$/kg)---------							
Mean	6.1	2.1	41	2.88	0.32	0.19	8.3	0.41	263.9	0.4	7.0
Std.	0.6	0.4	30	0.65	0.16	0.10	2.1	0.05	0.8	0.2	1.0
Optimum range[a]	6.5-7.5			50-80	0.41-1.32	0.33-0.50					

[a] Ranges were based on data in Buchholz (1983).

3. RESULTS AND DISCUSSION

A. Correlation, multiple linear regression, and path analysis

The Pearson correlation coefficients in Tables 3 and 4 show the overall linear relationship between two variables. Examination of the results shows that some variables (Mg vs. Ca, 0.98; Table 3) are more correlated than others (K vs. Ca, -0.13) and some variables have a positive correlation (P vs. yield, 0.27; Table 4) while others have a negative correlation (OM vs. EC_a, -0.15). As seen in Table 3, when soil nutrient levels increase, rice yield may not only increase (for example, with Ca or Mg), but may decrease (for example, with P_2O_5 and K). This situation is due to the interaction among variables limiting crop yield, and provides a motivation for using path analysis to obtain more insight about the correlation structure of the data.

Path analysis (Tables 5 and 6) gives a more detailed understanding of the relationships among limiting factors and their contribution to crop yield. Underlined diagonal elements in these Tables are partial regression coefficients from the multiple linear regression and indicate direct effects of each predictor variable on the dependent variable (crop yield in this case). Off-diagonal elements indicate the indirect effects of each variable on yield acting through the other, correlated variables.

Table 3 Correlation coefficients between variables for field 1.

	SPAD	pH	EC	OM	P_2O_5	Ca	K	Mg	Na	N	SiO_2
pH	0.35**										
EC	0.25*	0.22									
OM	-0.19+	0.08	-0.01								
P_2O_5	-0.39**	-0.32**	-0.31**	0.07							
Ca	0.58**	0.68**	0.49**	-0.02	-0.46**						
K	-0.11	-0.07	-0.03	0.03	0.25	-0.13					
Mg	0.59**	0.69**	0.51**	-0.02	-0.48**	0.98**	-0.14				
Na	0.39**	0.34**	0.26*	0.02	-0.16	0.50**	-0.11	0.43**			
N	-0.08	0.05	-0.01	0.32**	0.05	-0.05	-0.01	-0.06	-0.01		
SiO_2	0.54**	0.59**	0.45**	0.06	-0.39**	0.70**	0.03	0.69**	0.57**	0.03	
Yield	0.16	0.03	0.24*	-0.02	-0.14	0.19+	-0.15	0.19	0.02	-0.05	0.20+

+, *, ** significant at P<0.1, 0.05, 0.01 levels, respectively.

Table 4 Correlation coefficients between variables for field 2.

	pH	OM	P	Ca	Mg	K	CEC	EC_a	Elev	Slope
OM	0.13*									
P	0.21**	0.53**								
Ca	0.39**	0.55**	0.51**							
Mg	-0.23**	0.48**	0.44**	0.62**						
K	0.05	0.60**	0.85**	0.62**	0.64**					
CEC	-0.39**	0.47**	0.43**	0.65**	0.90**	0.65**				
EC_a	-0.12*	-0.15**	-0.21**	0.03	0.16**	-0.04	0.10*			
Elev	0.32**	0.45**	0.56**	0.41**	0.35**	0.56**	0.24**	-0.14*		
Slope	-0.33**	-0.18**	-0.26**	-0.04	0.16**	-0.16**	0.21**	0.29**	-0.43**	
Yield	0.08	0.09+	0.27**	0.03	0.06	0.13*	0.03	-0.45**	0.06	-0.24**

+, *, ** significant at P<0.1, 0.05, 0.01 levels, respectively.

Path analysis provided quantitative insights on "contribution paths" of predictor variables to a response variable. For example, the indirect contribution of Mg through Ca to rice yield was positive and relatively greater (0.51, Table 5) while that of Ca through Mg was negative and smaller (-0.37). Path analysis also enabled more appropriate selection of important factors limiting crop yield by considering cause-and-effect relationships among predictor and response variables. For example, although the simple correlations of Mg (r=0.06), CEC (r=0.03), and Ca (r=0.03) to corn yield were low and not significant for field 2 (Table 4), path analysis revealed a significant direct positive contribution of

Table 5 Path analysis direct effects (diagonal, underlined) and indirect effects of limiting factors for rice yield ($R^2 = 0.14^{+}$) and SPAD readings ($R^2 = 0.46^{**}$) in field 1.

Independent variable	Direct and indirect contribution to dependent variables							r
For rice yield	pH	EC	Ca	K	Mg	Na	SiO_2	
pH	-0.20	0.03	0.35	0.01	-0.26	-0.08	0.17	0.03
EC	-0.04	0.15	0.25	0.01	-0.20	-0.06	0.13	0.24^{*}
Ca	-0.13	0.07	0.52	0.02	-0.37	-0.12	0.20	0.19^{+}
K	0.01	-0.01	-0.07	-0.18^{+}	0.05	0.03	0.01	-0.15
Mg	-0.14	0.08	0.51	0.03	-0.38	-0.10	0.20	0.19
Na	-0.07	0.04	0.26	0.02	-0.17	-0.23^{+}	0.17	0.02
SiO_2	-0.12	0.07	0.36	-0.01	-0.26	-0.13	0.29^{+}	0.20^{+}
For SPAD readings	pH	EC	OM	P_2O_5	Mg	Na	SiO_2	
pH	-0.18	-0.03	-0.01	0.04	0.34	0.03	0.17	0.35^{**}
EC	-0.04	-0.16	0.01	0.04	0.25	0.03	0.13	0.25^{*}
OM	-0.02	-0.00	-0.18^{*}	-0.01	-0.01	-0.01	0.02	-0.19^{+}
P_2O_5	0.06	0.05	-0.01	-0.12	-0.23	-0.02	-0.11	-0.39^{**}
Mg	-0.12	-0.08	0.01	0.06	0.49	0.04	0.20	0.59^{**}
Na	-0.06	-0.04	-0.01	0.02	0.21	0.10	0.16	0.39^{**}
SiO_2	-0.11	-0.07	-0.01	0.05	0.34	0.06	0.28^{*}	0.54^{**}

$^{+}, ^{*}, ^{**}$ significant at P<0.1, 0.05, 0.01 levels, respectively.

Table 6 Path analysis direct effects (diagonal, underlined) and indirect effects of limiting factors on yield for field 2 ($R^2=0.31^{*}$).

Independent variable	Direct and indirect contribution to crop yield									r
	pH	OM	P	Ca	Mg	K	CEC	EC_a	Elev	Slope	
pH	0.37^{*}	-0.01	0.06	-0.17	-0.07	-0.01	-0.12	0.05	-0.09	0.07	0.08
OM	0.05	-0.07	0.16	-0.24	0.14	-0.07	0.15	0.06	-0.13	0.04	0.09^{+}
P	0.08	-0.04	0.31^{**}	-0.22	0.13	-0.10	0.14	0.08	-0.16	0.05	0.27^{**}
Ca	0.14	-0.04	0.16	-0.43^{*}	0.18	-0.08	0.21	-0.01	-0.11	0.01	0.03
Mg	-0.08	-0.03	0.14	-0.27	0.29^{*}	-0.08	0.29	-0.06	-0.10	-0.03	0.06
K	0.02	-0.04	0.26	-0.27	0.19	-0.12	0.21	0.02	-0.16	0.03	0.13^{*}
CEC	-0.14	-0.03	0.13	-0.28	0.26	-0.08	0.32	0.04	-0.07	-0.04	0.03
EC_a	-0.04	0.01	-0.06	-0.01	0.05	0.01	0.03	-0.40^{**}	0.04	-0.06	-0.45^{**}
Elev	0.12	-0.03	0.17	-0.18	0.10	-0.07	0.08	0.05	-0.28^{**}	0.09	0.06
Slope	-0.12	0.01	-0.08	0.02	0.05	0.02	0.07	-0.12	0.12	-0.20^{**}	-0.24^{**}

$^{+}, ^{*}, ^{**}$ significant at P<0.1, 0.05, 0.01 levels, respectively.

Mg (0.29, $P<0.1$, Table 6) and a negative contribution of Ca (-0.43, $P<0.1$). A considerable indirect positive contribution of Mg through CEC (0.29, Table 6) and a negative contribution of Mg through Ca (-0.27) were also found. Based on the correlation analysis alone (Table 4), increasing the application of P and/or K would increase corn yield. With the help of path analysis, however, one might decide that corn yield would also increase with added Mg but would decrease with added Ca. The need for additional Mg is consistent with mean Mg levels being below the optimum range (Table 2). The negative Ca-yield relationship is not because excess Ca reduces corn yield. Rather, higher Ca levels are an indication of higher clay content in the soil since Ca is a major part of CEC. Soils with higher clay contents have lower plant-available water holding capacity, which could be expected to decrease corn yield in dry years such as 1997.

B. Rice Growth and Limiting Factors

From the results for the rice paddy field (Table 5), the coefficient of determination for estimating rice yield (0.14) was lower than that for chlorophyll content (SPAD, 0.46). Also, the correlation coefficient between yield and chlorophyll content was low (0.16, Table 3). This was possibly due to heavy rain during the summer rainy season of 1999 and pest damage before harvest. Using multiple regression with the forward variable selection option, pH, EC, Mg, Na, and SiO_2 were selected as important variables explaining both rice yield and growth (SPAD readings), while Ca and K were selected only for rice yield and OM and P_2O_5 were selected only for SPAD readings. The other variables did not meet the significance level for entry into the model (SLENTRY=0.5), and were not included in the path analysis (Table 5).

Correlation analysis (Table 3) showed that EC (0.24, $P<0.05$), Ca (0.19, $P<0.1$) and SiO_2 (0.20, $P<0.1$) had significant positive correlations and K (-0.15) had a negative correlation with rice yield. For the SPAD chlorophyll reading, Ca (0.58, $P<0.01$), Mg (0.59, $P<0.01$), Na (0.39, $P<0.01$) and SiO_2 (0.54, $P<0.01$) had significant positive correlations and P_2O_5 (-0.39, $P<0.01$) had a significant negative correlation. Based on correlation results, Ca and SiO_2 were nutrients that should be increased to improve rice growth and yield, commensurate with the fact that field means were below the recommended range (Table 1). Increases in K and P_2O_5 would not increase rice yield, and field means for these nutrients were above the recommended range (Table 1). The negative relationships of K and P_2O_5 to rice growth might be due to over-supply of the nutrients from past field management.

Path analysis (Table 5) provided additional information about the contribution paths of soil variables to rice yield and growth. For example, Ca had the highest direct effect (0.52) and indirect effect via Mg (-0.37) on rice yield. The direct effect of SiO_2 on yield and the indirect effect through Ca were 0.29 ($P<0.1$) and 0.36, respectively. The indirect effect of Mg through Ca (0.51) was higher than the direct effect (-0.38). For pH, indirect effects through Ca (0.35) and Mg (-0.26) were higher than the direct effect. Mg and Na were found to be important as they produced considerable indirect effects on rice yield. Reduction of Na would be a reasonable decision that could possibly increase rice yield for the field since it generally showed a negative indirect effect through most of the other variables. Although an optimum range for Na is not established (Table 1), lower Na levels are generally assumed to give better yields in Korean rice production.

Most measured variables were significantly correlated with SPAD chlorophyll reading, in the following order: Mg (0.59, $P<0.01$), SiO_2 (0.54, $P<0.01$), P_2O_5 (-0.39, $P<0.01$), Na (0.39, $P<0.01$) and pH (0.35, $P<0.01$). Three variables, pH, EC, and SiO_2 exhibited indirect effects of 0.34, 0.25 and 0.34 through Mg to SPAD readings, respectively. Although pH showed a positive correlation ($r=0.35$, $P<0.01$) with SPAD readings, it should be noted (1) that the overall relationship was mainly due to the indirect positive effects acting through Mg and SiO_2, and (2) that pH not only showed negative direct effects, but also negatively affects indirect effects of other variables on SPAD readings. Therefore, an increase in pH might not be a reasonable option for increasing rice growth on this field, where mean pH was only slightly below the optimum range (Table 1). This is an example where path analysis provided additional information that could lead to a different management decision than the more conventional correlation analysis.

C. Comparing Small Rice Paddy Field and Large Upland Corn Field

Factors limiting crop growth and yield may be different for different fields due to reasons such as soils, crop variety, weather, and field management history. The rice paddy field (field 1) was small in size (0.3 ha), flat, and mostly flooded by irrigation and heavy rain. Therefore, the amount and direction of water flow had a large effect on nutrient distribution (Chung et al., 2000). EC, Ca, and SiO_2 showed significant correlations with both rice growth and yield for field 1 (Table 3). Topographic variables such as elevation and/or slope play considerable roles in the large (36 ha) upland field (field 2). P, K, EC_a, and slope showed significant correlations with corn yield for field 2 (Table 4). Although correlation coefficients for elevation and slope were relatively low, they have significant direct effects ($P<0.01$, Table 6) on corn yield and significant correlations ($P<0.01$, Table 4) with many other limiting factors.

Site variables behaved differently in the two fields. Between the fields, direct contributions were of opposite sign for pH, Mg, and Ca, while they were of the same sign for K. For example, pH had a negative direct contribution to rice yield in field 1 but a positive direct contribution to corn yield in field 2. The coefficient of determination for corn yield was greater than that for rice yield. These phenomena, however, were based only on one site-year of data. To generalize these results, data for multiple years should be investigated. Also, knowledge of other factors such as climatic conditions, previous crop history, and pest damage would be needed to establish a more complete model of yield response to limiting factors.

D. Potential for Use of Path Analysis in Precision Agriculture

Although this study considered a number of variables, the regressions developed were not highly predictive. From this we can infer that there are many other yield-limiting factors not included in the study. Those factors may include other soil and plant parameters, management factors, pest information, and climatic conditions. Prediction of crop yield could be improved using nonlinear regression techniques. Previous work (e.g., Drummond et al., 2003; Sudduth et al., 1996) using similar data sets from field 2 showed that nonlinear techniques such as neural networks (NN) and projection pursuit regression (PPR) produced higher coefficients of determination and lower errors of prediction. For 1993 corn yield, coefficient of determination and standard error of prediction (SEP) were 0.13 and 0.65 Mg/ha with a stepwise multiple linear regression (SMLR), and were 0.51 and 0.48 Mg/ha with a PPR (Sudduth et al., 1996). For 1997 corn yield, the same data used in this paper, the coefficient of determination and SEP were 0.31 and 0.71 Mg/ha using a forward variable selection option in this study, while SEP values were 0.84, 0.70, and 0.69 Mg/ha with SMLR, PPR, and NN methods, respectively, in a previous study (Drummond et al., 2003). The PPR and NN approaches focus on yield prediction and use the concept of a transformed or "latent variable". Therefore understanding the contributions of the original variables and interrelationships among them is difficult. Also these methods require a significantly greater number of observations than do linear methods.

Another reason for poor predictive performance in this study might be nonlinearity of response curves and nonlinear interrelationships among predictor variables. While path analysis based on correlation analysis and partial linear regression enables us to obtain insight into the correlation structure by partitioning the overall relationship into direct and indirect effects, it assumes linear effects of the limiting factors on crop yield.

Solving these problems may require the use of nonlinear response functions. Nutrient response curves described by Tisdale et al. (1999) and/or the boundary line approach demonstrated by Kitchen et al. (1999) may be used to transform nonlinear data into data linearly related to crop yield. To help understand the results of path analysis, other methods such as principal component analysis (PCA) could be used. PCA generates new decorrelated data sets (principal components, PCs) from the original correlated data sets by grouping the most highly correlated variables. Examination of the PCs may also provide some information on interrelationships among the original data sets (e.g., Chung et al., 2001).

4. CONCLUSIONS

Path analysis was applied to investigate yield-limiting factors and their interactions with one another for two research sites, a small rice paddy field in Korea and a large upland corn field in Missouri, USA. The results provided more information on the direct and indirect effects of limiting factors on crop yield than commonly used methods such as correlation analysis and multiple linear regression. In general, results agreed with published optimum ranges of nutrients for rice and corn production.

From correlation analysis of the rice field data, EC, Ca, and SiO_2 had significant correlations with rice yield, while pH, Ca, Mg, Na, SiO_2, and P_2O_5 had significant correlations with the SPAD chlorophyll reading. Path analysis provided additional information about the importance and contribution paths of soil variables to rice yield and growth. Path analysis also enabled more appropriate selection of important factors limiting crop yield by considering cause-and-effect relationships among predictor and response variables. For example, although pH showed a positive correlation ($r=0.35$) with SPAD readings, the correlation was mainly due to the indirect positive effects acting through Mg and SiO_2. The direct effects of pH on SPAD were negative, so an increase in pH might not increase rice growth.

Overall, the coefficient of determination for rice yield (0.14) was lower than that for SPAD chlorophyll reading (0.46), and the correlation coefficient between yield and SPAD reading was low (0.16). This was attributed to heavy rain during the summer rainy season and pest damage before harvest. Therefore, other factors such as climatic conditions and pest damage would be needed to establish a more complete model of yield response to limiting factors.

For the large upland Missouri corn field, topographic factors including elevation and slope had significant direct effects ($P<0.1$) on yield and highly significant correlations ($P<0.01$) with other limiting factors. Based on correlation analysis, increases in P and K would increase corn yield. With the help of path analysis, increased Mg levels were also indicated for improving corn yield

Path analysis assumes a linear relationship of limiting factors to crop yield response, but the true relationships are non-linear in most cases. Results of path analysis could be improved if the non-linear responses were linearized using tools such as the boundary line method and/or general nutrient response models.

ACKNOWLEDGEMENTS

This research was part of an international joint project, "Development of Precision Farming for Sustainable Agricultural Production", between the Rural Development Administration, Republic of Korea and the University of Missouri, USA, from 1999 to 2001. The authors appreciate the contributions of Drs. Yoo-Hak Kim (NIAST, RDA) and Newell Kitchen (USDA-ARS) to interpretation of analysis results.

REFERENCES

1. Birrell, S. J., K. A. Sudduth and S. C. Borgelt. 1996. Crop yield mapping: comparison of yield monitors and mapping techniques. Computers and Electronics in Agriculture 14: 215-233.
2. Buchholz, D. D. 1983. Soil Test Interpretations and Recommendations Handbook. University of Missouri, College of Agriculture, Department of Agronomy, Columbia, Mo.
3. Cambardella, C. A., T. S. Colvin, D. L. Karlen, S. D. Logsdon, E. C. Berry, J. K. Radke, T. C. Kaspar, T. B. Parkin and D. B. Jaynes. 1996. Soil property contributions to yield variation patterns. In Proceedings of the Third International Conference on Precision Agriculture, P. C. Robert et al. (eds.), pp. 189-195. Madison, Wisc.: ASA, CSSA, and SSSA.
4. Chung, S. O., K. A. Sudduth, S. T. Drummond and N. R. Kitchen. 2001. Analysis of soil, topographic and yield data using nested variograms. In Proceedings of the Third International Conference on Geospatial Information in Agriculture and Forestry, B. J. Petoskey (ed.), Unpaginated CD-ROM. Ann Arbor, Mich.: Veridian.
5. Chung, S. O., J. H. Sung, K. A. Sudduth, S. T. Drummond and B. K. Hyun. 2000. Spatial variability of yield, chlorophyll content, and soil properties in a Korean rice paddy field. In proceedings of Fifth International Conference on Precision Agriculture, P. C. Robert et al. (eds.), Unpaginated CD-ROM. Madison, Wisc.: ASA, CSSA, and SSSA.
6. Dobermann, A. 1994. Factors causing field variation of

direct-seeded flooded rice. Geoderma 62:125-150.

7. Drummond, S. T., K. A. Sudduth, A. Joshi, S. J. Birrell and N. R. Kitchen. 2003. Statistical and neural methods for site-specific yield prediction. Transactions of the ASAE 46(1):5-14.
8. Fraisse, C. W., K. A. Sudduth and N. R. Kitchen. 2001a. Calibration of the CERES-MAIZE model for simulating site-specific crop development and yield on claypan soils. Applied Engineering in Agriculture 17(4):547-556.
9. Fraisse, C. W., K. A. Sudduth and N. R. Kitchen. 2001b. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. Transactions of the ASAE 44(1):155-166.
10. Gravois, K. A. and R. S. Helms. 1992. Path analysis of rice yield and yield components as affected by seeding rate. Agronomy Journal 84:1-4.
11. Institute of Agricultural Science. 1984. Official Soil Series Description: Soil Survey Material No. 10. Institute of Agricultural Science, Rural Development Administration, Suwon, Republic of Korea.
12. Irmak, A., J. W. Jones, W. D. Batchelor and J. O. Paz. 2001. Estimating spatially variable soil properties for application of crop models in precision farming. Transactions of the ASAE 44(5):1343-1353.
13. Jensen, J. R. 1996. Introductory digital image processing: a remote sensing perspective. 2nd ed. Upper Saddle River, N.J.: Prentice-Hall, Inc.
14. Kasap, Y., S. Irmak and H. Gunal. 1999. Path analysis of some heavy metal adsorption by soil. In Proceedings of the Fourth International Conference on Precision Agriculture, P. C. Robert et al. (eds.), pp. 91-100. Madison, Wisc.: ASA, CSSA, and SSSA.
15. Kitchen, N. R., K. A. Sudduth and S. T. Drummond. 1999. Soil electrical conductivity as a crop productivity measure for claypan soils. Journal of Production Agriculture 12: 607-617.
16. Krishnasamy, R., and K. K. Mathan. 2001. Path coefficient analysis of zinc and boron adsorption in soils. Communications in Soil Science and Plant Analysis 32(3&4):465-475.
17. Mallarino, A. P., P. N. Hinz and E. S. Oyarzabal. 1996. Multivariate analysis as a tool for interpreting relationships between site variables and crop yields. In Proceedings of the Third International Conference on Precision Agriculture, P. C. Robert et al. (eds.), pp. 151-158. Madison, Wisc.: ASA, CSSA, and SSSA.
18. NIAST. 1999. Criteria for crop fertilization. Technical Report, National Institute of Agricultural Science and Technology, Rural Development Administration, Suwon, Republic of Korea.
19. Pantone, D. J., J. B. Baker and P. W. Jordan. 1992. Path analysis of red rice (*Oryza sativa* L.) competition with cultivated rice. Weed Science 40:313-319.
20. Ssango, F., P. R. Speijer, D. L. Coyne and D. D. Waele. 2004. Path analysis: a novel approach to determine the contribution of nematode damage to East African Highland banana (*Musa* spp., AAA) yield loss under two crop management practices in Uganda. Field Crops Research (in press).
21. Sudduth, K. A. 1998. Engineering for precision agriculture – Past accomplishments and future directions. In Agricultural Machine Systems, pp. 29-36, SAE Spec. Pub. 1383. Warrendale, Pa.: SAE.
22. Sudduth, K. A., S. T. Drummond, S. J. Birrell and N. R. Kitchen. 1996. Analysis of spatial factors influencing crop yield. In Proceedings of the Third International Conference on Precision Agriculture, P. C. Robert et al. (eds.), pp. 129-140. Madison, Wisc.: ASA, CSSA, and SSSA.
23. Sudduth, K. A., S. T. Drummond and N. R. Kitchen. 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. Computers and Electronics in Agriculture 31:239-264.
24. Sudduth, K. A., N. R. Kitchen, G. A. Bollero, D. G. Bullock and W. J. Wiebold. 2003. Comparison of electromagnetic induction and direct sensing of soil electrical conductivity. Agronomy Journal 95(3):472-482.
25. Sung, J. H., S. O. Chung and S. Y. Hong. 1999. Site-specific rice crop fertilizing based on chlorophyll content: describing chlorophyll content variation. In Proceedings of International Conference on Agricultural Engineering, Z. Dechao et al. (eds.), pp. 104-108. Beijing, China: China Agricultural University Press.
26. Tisdale, S. L., J. L. Havlin, J. D. Beaton and W. L. Nelson. 1999. Soil Fertility and Fertilizers: An Introduction to Nutrient Management. 6th ed. Upper Saddle River, N.J.: Prentice-Hall, Inc.
27. Webb, R. A. 1972. Use of boundary line in the analysis of biological data. Journal of Horticultural Science 47:309-319.
28. Williams, W. A., M. B. Jones and M. W. Demment. 1990. A concise table for path analysis statistics. Agronomy Journal 82:1022-1024.