SPATIAL ANALYSIS OF YIELD-LIMITING FACTORS

K. A. Sudduth, S. T. Drummond, S.J. Birrell, and N.R. Kitchen[1]

ABSTRACT

The spatial relationship between crop yields and soil and site parameters was modeled using several methods. Yield maps estimated by projection pursuit regression and neural network analysis agreed well with measured yields. These methods also allowed generation of response curves for estimated yield as a function of each of the input parameters. These response curves were useful for investigating the relationship between yields and individual soil and site parameters.

BACKGROUND

Understanding the functional relationship of crop yield to other spatial factors is a basic need for successful site-specific crop management (SSCM). A first approximation to this relationship can be obtained with conventional nutrient recommendation procedures (for example, Buchholz, 1983). However, these recommendation procedures are generally based upon response data averaged over a large geographic area, thus diluting the precision of the response relationship. To apply inputs with the precision needed for SSCM, it could be more appropriate to develop individual response functions for particular soils or soil associations, or perhaps even for a particular field, or for similar areas within a field.

Another shortcoming of the current nutrient recommendation procedures is they necessarily assume that all factors limiting yield are included in the recommendation process. When the procedures are applied on a point-by-point basis within a field, there may be areas in which crop growth and yield are limited by other factors, such as water availability. In these portions of the field, the current recommendation procedures will not accurately relate crop yield to spatial soil and site parameters.

The use of analysis techniques to predict yield from input parameters is of importance in developing SSCM methods and recommendation procedures, but even more important is the ability to use the techniques and models to understand yield response (or sensitivity) to changes in critical factors. One approach to developing such an understanding of these relationships is through the application of crop growth models (Hoogenboom et al., 1993). Another approach is based on empirical analysis of multivariate spatial data. We previously used the empirical approach to study the spatial relationship between soil properties and yields for a research field in central Missouri

---

[1] Agricultural Engineer, Cropping Systems and Water Quality Research Unit, USDA-Agricultural Research Service, Columbia, MO; Scientific Programmer/Analyst, Dept. of Biological and Agricultural Engineering; Research Assistant Professor, Dept. of Biological and Agricultural Engineering; and Research Assistant Professor, Dept. of Soil and Atmospheric Sciences; University of Missouri, Columbia, MO 65211. Presented at the 26th North Central Extension-Industry Soil Fertility Conference, November 20-21, 1996, St. Louis, MO.

(Drummond et al., 1995). For this paper, we have expanded on the previous study and applied this approach to additional datasets.

OBJECTIVES

The overall objective was to study the relationship between yields and soil properties (i.e. nutrient availability, organic matter) and site properties (i.e. elevation) on a spatial basis. Specific objectives were to: (1) develop and/or evaluate methods for predicting spatial crop yields, and (2) use the model results to investigate the sensitivity of crop yield to variations in soil and site parameters.

MATERIALS AND METHODS

Data were collected on two fields, 90 acres and 70 acres in size, located near Centralia, in central Missouri. The soils of the area are characterized as claypan soils, primarily of the Mexico-Putnam association (fine, montmorillonitic, mesic Udollic Ochraqualfs). These soils are poorly drained and have a restrictive, high-clay layer (the claypan) occurring below the topsoil. The two fields were managed in a high yield goal, high input, minimum till corn-soybean rotation. Fertilizer and chemical inputs were applied at a single rate.

Data were obtained for one field (Field 1) in 1993 (corn), 1994 (soybean), and 1995 (grain sorghum). Grain sorghum was planted in this field in 1995, rather than corn, because an excessively wet spring delayed planting until mid-June. Yield data for the other field (Field 2) were obtained only in 1995 (soybean). Conditions for crop production were quite different between the three years. The 1993 growing season was characterized by heavy and frequent rains, with an annual precipitation of 62 in. Yield reductions were observed in lower portions of the landscape, due to excess water. The 1994 precipitation of 32 in. was only slightly below average, but less than 2 in. of rainfall was received in July and August and crops experienced drought stress during much of the growing season. In 1995, precipitation was 45 in., with an excessively wet planting season which again caused stand problems and some yield reductions in the lower portion of the landscape.

Data Acquisition

Data obtained on the study fields included grain yield, elevation, and a number of soil properties. Grain yield measurements were obtained using a full-size combine equipped with a commercial yield sensing system and global positioning system (GPS) receiver, using data collection and processing techniques described by Birrell et al. (1996). Detailed topographic data were obtained on each field, using a total station surveying instrument and standard mapping procedures.

Based on our previous work (Sudduth et al., 1995), topsoil depth above the claypan was estimated from soil conductivity. A mobile measurement system described by Kitchen et al. (1996) was used to obtain root-zone soil conductivity data with a commercial electromagnetic induction (EM) sensor. The actual depth of topsoil was measured at a set of randomly selected calibration points and a regression between topsoil depth and the inverse of soil conductivity was developed (Field 1: $r^2 = 0.90$, std. err. = 2.6 in.; Field 2: $r^2 = 0.89$, std. err. = 3.7 in.). These regressions were then applied to convert the EM data to topsoil depth.

Field 1 was soil sampled on a 98 ft (30 m) grid in the spring of 1995. A hand soil probe was used to collect soil cores to a depth of 8 in. Three soil cores obtained within a 3.3 ft (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 phosphorus, potassium, pH, organic matter, calcium and magnesium. Cation exchange capacity (CEC) and magnesium saturation were calculated according to standard procedures (Buchholz, 1983). Field 2 was soil sampled on a 82 ft (25 m) grid in the spring of 1996. Procedures were identical to those for Field 1, except that 8 cores were combined at each sample position.

Data Analysis

Yield and topsoil depth data were analyzed using geostatistics, and appropriate semivariogram models and parameters were used to krige the data to a grid with a 33 ft (10 m) cell size. Data from the grid cell centered closest to each soil sampling point was extracted and combined with the soil sample data for analysis. If any data was missing for a grid cell, that cell was eliminated from the analysis. The whole-field datasets ranged in size from 301 to 436 observations (Table 1).

Additional datasets were created for analysis by dividing each field into 5 sub-field areas on the basis of elevation and topsoil depth. It was thought that the relationship of yield to soil and site parameters might be more predictable within these areas than across the entire field. The two relatively static parameters of elevation and topsoil depth were chosen because previous analysis indicated that these had the most consistent impact on yields of all the measured parameters in the dataset. To create the sub-field areas, each field was first divided into areas of low (<10 in.), medium, and high (>20 in.) topsoil depth. The medium and high topsoil depth areas were then subdivided into the lower 1/3 of the landscape and the higher 2/3 of the landscape (Figure 1). The sub-field datasets ranged in size from 14 to 232 observations (Table 1).

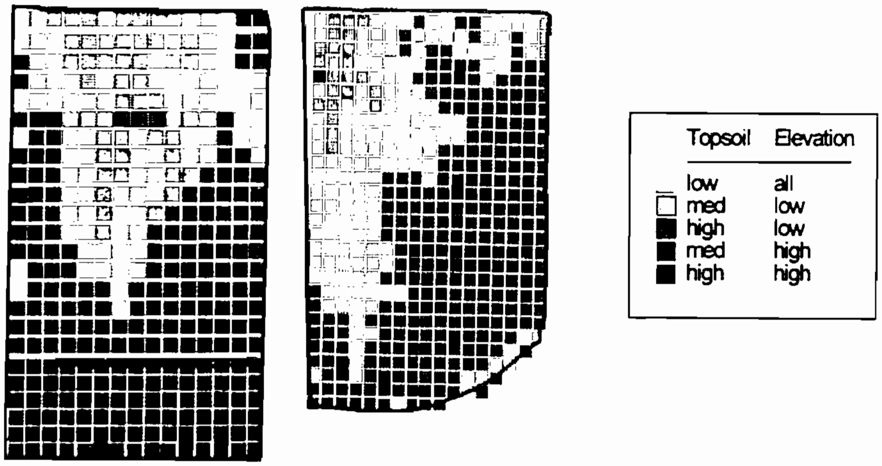

Figure 1. Sub-field areas classified by topsoil depth and elevation.

Table 1. Correlations between yields and soil and site properties for whole-field and sub-field areas.

Field/ Year	Area	# Obs	P	K	pH	Organic Matter	CEC	Ca	Mg	Mg Sat.	Topsoil Depth	Elevation	Slope	Curvature
Field1 1993	all	318	-0.08	-0.01	0.07	0.06	-0.15*	-0.04	-0.17*	-0.16*	-0.13*	-0.01	-0.05	-0.08
	lt	16	-0.46*	-0.74*	0.64*	-0.27	-0.60*	-0.20	-0.45*	-0.21	-0.28	0.48*	-0.06	-0.25
	mtle	67	0.14	0.11	0.44*	0.31*	-0.25*	0.07	-0.18	-0.05	0.06	0.14	-0.22*	-0.08
	htle	43	-0.03	-0.10	-0.26*	0.08	0.07	-0.18	-0.12	-0.21	0.15	0.04	0.08	0.08
	mthe	170	-0.09	-0.03	-0.01	-0.02	-0.16*	-0.12	-0.21*	-0.25*	-0.10	-0.14*	-0.04	-0.09
	hthe	22	-0.61*	-0.10	-0.31	-0.22	-0.31	-0.33	-0.67*	-0.77*	-0.41*	-0.47*	-0.33	0.17
Field1 1994	all	344	0.09	-0.05	0.13*	-0.14*	-0.16*	-0.08	-0.12*	-0.10*	0.39*	-0.22*	-0.10*	0.11*
	lt	17	0.45*	0.41*	-0.57*	0.03	0.13	-0.25	-0.08	-0.27	0.13	-0.24	-0.10	-0.09
	mtle	71	0.57*	0.31*	0.27*	0.14	-0.06	0.00	0.07	0.23*	0.08	0.03	0.01	0.14
	htle	47	0.40*	-0.08	0.10	-0.10	0.12	0.27*	0.33*	0.43*	0.52*	-0.65*	-0.07	0.01
	mthe	187	0.11	0.03	0.33*	-0.17*	-0.25*	-0.03	-0.21*	-0.21*	0.07	-0.02	-0.22*	0.14*
	hthe	22	0.17	-0.20	0.31	0.36*	-0.34	0.02	-0.09	0.08	0.09	0.52*	-0.40*	-0.40*
Field1 1995	all	301	0.06	-0.02	0.05	0.16*	-0.06	-0.07	-0.05	0.01	-0.03	0.26*	-0.09	-0.06
	lt	14	-0.28	-0.31	0.17	0.16	-0.20	-0.18	-0.09	0.08	-0.03	0.08	-0.01	-0.02
	mtle	72	-0.47*	-0.55*	-0.04	-0.32*	-0.44*	-0.61*	-0.36*	-0.16	0.14	-0.03	-0.22*	-0.19
	htle	40	-0.44*	-0.38*	-0.16	-0.11	-0.35*	-0.62*	-0.39*	-0.27*	-0.36*	0.34*	0.48*	-0.17
	mthe	161	-0.00	-0.05	0.03	0.18*	0.07	0.06	0.01	0.04	0.05	0.24*	-0.10	-0.01
	hthe	21	-0.01	-0.44*	-0.01	0.32	-0.11	-0.12	-0.06	0.01	0.12	0.11	-0.10	-0.54*
Field2 1995	all	436	0.05	-0.37*	-0.10*	-0.43*	-0.48*	-0.52*	-0.44*	-0.34*	-0.09*	0.49*	-0.08*	-0.16*
	lt	68	0.19	-0.28*	-0.12	-0.25*	-0.42*	-0.49*	-0.35*	-0.24*	0.27*	0.57*	-0.25*	-0.18
	mtle	81	-0.52*	-0.28*	0.06	-0.21*	-0.11	-0.02	-0.10	-0.01	-0.33*	0.48*	0.52*	-0.29*
	htle	26	-0.38*	-0.42*	0.04	-0.21	-0.32	-0.27	-0.34*	-0.28	-0.44*	-0.21	-0.41*	-0.22
	mthe	232	-0.05	-0.05	-0.24*	-0.09	-0.09	-0.24*	-0.01	0.11	0.37*	-0.02	0.05	-0.15*
	hthe	29	-0.01	-0.76*	-0.22	-0.56*	-0.66*	-0.73*	-0.61*	-0.52*	-0.62*	-0.43*	0.27	0.18

Area designations: lt = low topsoil depth; mtle = medium topsoil, low elevation; htle = high topsoil, low elevation; mthe = medium topsoil, high elevation; hthe=high topsoil, high elevation. Starred correlations are significant at the 0.10 level.

Standard correlation and stepwise multiple linear regression (SMLR) analyses were completed on each whole-field and sub-field dataset. Another analysis method used was projection pursuit regression (PPR). This nonparametric regression method requires only a few general assumptions about the shape of the regression surface, in contrast to the linearity constraints of SMLR (Friedman and Stuetzle, 1981). In PPR, the regression response (yield in this case) is modeled as the sum of a set of general (nonlinear) smooth functions of linear combinations of the independent (soil and site property) variables. In the SMLR and PPR analyses, yield data for each field-year were regressed against seven soil and site variables – phosphorus, potassium, pH, organic matter, topsoil depth, CEC, and elevation. The other original variables were not used in this analysis to reduce problems associated with colinearity. On these fields, we found that calcium, magnesium, and magnesium saturation were highly correlated with CEC, and slope was correlated with topsoil depth.

A neural network analysis was also used to model the data. Neural networks are computing systems composed of simple, highly interconnected processing units which respond to the sum of inputs from all connections in accordance with an activation function (Hopfield and Tank, 1986). For this study, a backpropagation network (BPN) with a sigmoid activation function was chosen. With this nonlinear activation function, the learning process for the network was nonlinear and nonparametric. The BPN was designed with three layers; an input layer consisting of seven input neurons, a hidden layer consisting of ten hidden neurons, and an output layer consisting of a single output neuron. Each layer was fully connected with the following layer in a feed-forward arrangement. Relatively few hidden neurons were used since this allowed training to be completed in a reasonable time frame, and also because a smaller number of neurons tends to limit the ability of the network to overfit the data. The dataset was randomly divided into training and testing sets for cross-validation as a further means to guard against overfitting. The BPN was trained with each field-year dataset in a separate session using the same seven input variables as were used in the SMLR and PPR analyses.

RESULTS AND DISCUSSION

Yield patterns for Field 1 varied considerably from year to year. Visual comparison of yield maps and soils maps from this field allowed us to associate some, but not all, yield patterns with soil variations. For example, the spatial pattern of 1994 yields showed some similarity to the pattern of topsoil depth variation across the field (Drummond et al., 1995).

Whole-field statistical correlations of yields to soil and site parameters (Table 1) were difficult to interpret in a way that yielded meaningful information. We concluded that these problems were likely due to a complex and nonlinear functional relationship between yield and soil properties, which was not well represented by the correlation statistic. Also, the form of the yield response function was likely different from region to region within the field, due to different factors controlling the expression of yield. In some cases, correlations calculated on a sub-field basis were more significant than those calculated on a whole-field basis (Table 1). However, this was not always the case, and it was apparent that linear correlation analysis was not the best approach to discerning the functional relationships between yield and soil and site properties.

Regression and Neural Network Analysis

SMLR analysis was applied to the yield and soil datasets, both for the entire field and for the five sub-field areas defined above. The best significant models selected for each field included from 4 to 6 variables, with r^2 values ranging from 0.13 to 0.43 (Table 2). Elevation was the only model variable common to all four datasets. Topsoil depth, organic matter, and phosphorus were also significant variables in a majority of cases. Although the SMLR analysis gave some insight into the relationships between yield and soil properties, the linearity constraints imposed by this model meant that it could not accurately describe the data.

Table 2. Model statistics for estimation of yield data as a function of seven soil and site parameters, using different analysis methods.

	Field 1 1993 (corn)	Field 1 1994 (soybean)	Field 1 1995 (sorghum)	Field 2 1995 (soybean)
Stepwise multiple linear regression (SMLR)				
r^2	0.13	0.25	0.15	0.43
std. error, bu/A	10.3	3.7	8.9	6.1
Projection pursuit regression (PPR)				
r^2	0.51	0.68	0.56	0.77
std. error, bu/A	7.6	2.4	6.4	3.8
Backpropagation neural network (BPN) analysis				
r^2	0.12	0.55	0.20	0.52
std. error, bu/A	10.2	2.8	8.4	5.6
Calibration dataset				
number of data points	318	344	301	436
mean yield, bu/A	116.0	24.2	77.9	32.6

Nonparametric regression analysis by PPR provided significantly better estimates of yield than did SMLR analysis using the same input datasets (Table 2). Estimated yield maps based on PPR compared well with actual yield maps (Figure 2). The best PPR estimations were obtained for field-years with well-defined, relatively large-scale yield patterns (Table 2, Figure 2). For example, the areas of highest yield for Field 1 in 1994 were reproduced well, as were the areas of lowest yield for Field 2 in 1995. The correspondence of PPR estimates and actual yields was weaker when the spatial structure of actual yield was less well-defined, as was the case for Field 1 in 1993 and 1995.

The BPN approach also showed promise for estimating yields (Table 2). Predicted yield maps were similar to actual yield maps, and to the maps predicted by PPR for Field 1 in 1994 and Field 2 in 1995 (Figure 2). The BPN maps reproduced the same major features as did the PPR maps, although at a lower level of accuracy. In general, the BPN-derived maps showed less localized variation in yield than those created with PPR. Yields estimated by BPN for Field 1 in 1993 and 1995 were less accurate. Some improvement in BPN yield estimates might be possible with additional refinement of network training parameters and procedures.

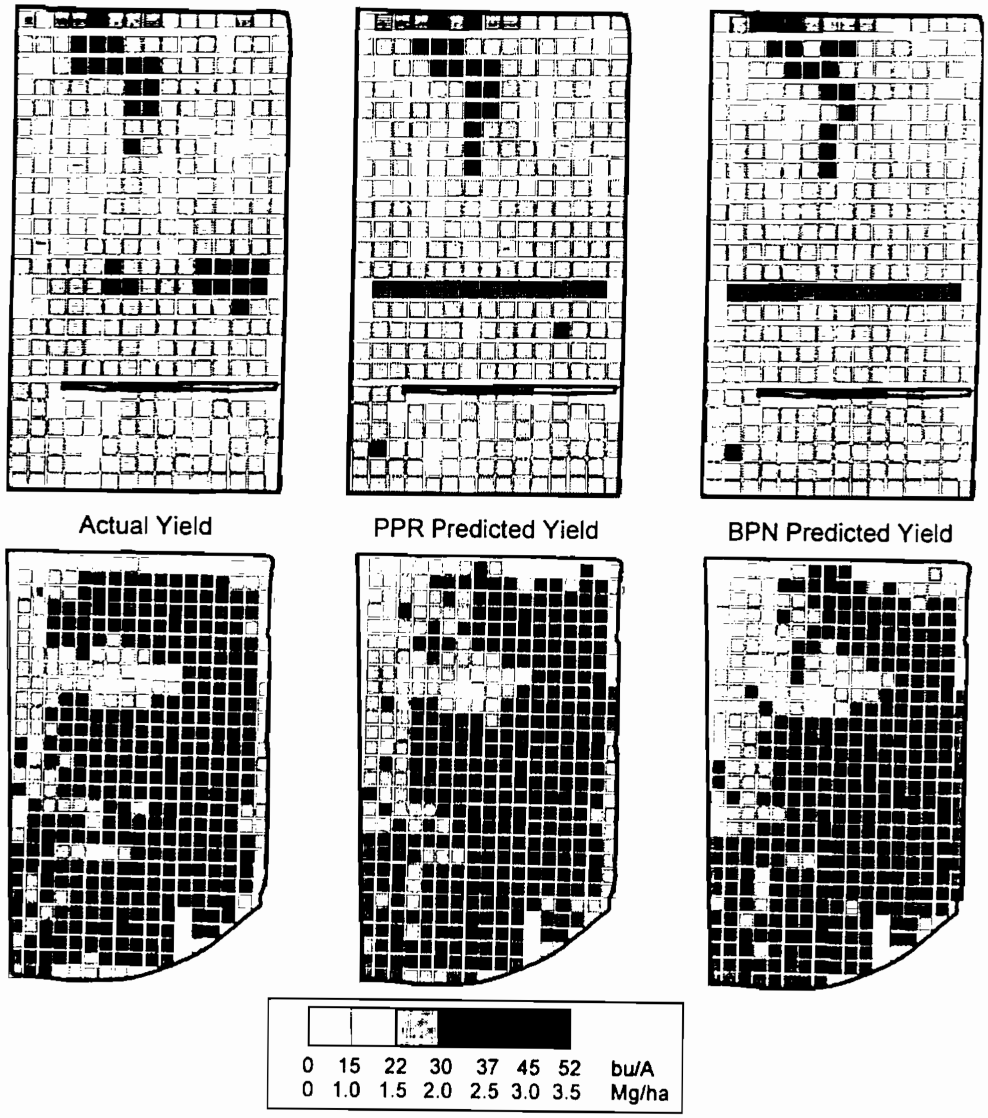

Figure 2. Measured, PPR predicted, and BPN predicted soybean yields for Field 1 in 1994 (top) and Field 2 in 1995 (bottom).

Yield Response Curves

The response of the PPR and BPN yield estimates to variations in the input parameters was investigated on a point-by-point basis. Sensitivity analyses were conducted by holding all but one of the model input parameters constant and varying the other parameter from a minimum to a maximum value. All response curves for each sub-field area were then combined into a mean response curve for that area. For generation of the response curves, all variables were normalized to a field-year mean of zero and unity standard deviation. This facilitated comparison of yield responses to the different soil and site parameters.

Response curves generated by the PPR and BPN analyses were similar overall. For example, the general response to elevation for the 1995 Field 2 yield data was positive for both methods (Figure 3). For this and most other parameters, the PPR curves exhibited more local variation than did the BPN curves. Since r^2 values for the PPR analyses were higher than those for the BPN (Table 2), this may reflect better modeling of actual trends in the data. On the other hand, further investigation of the results is needed to be sure that the PPR analysis is not overfitting the data.

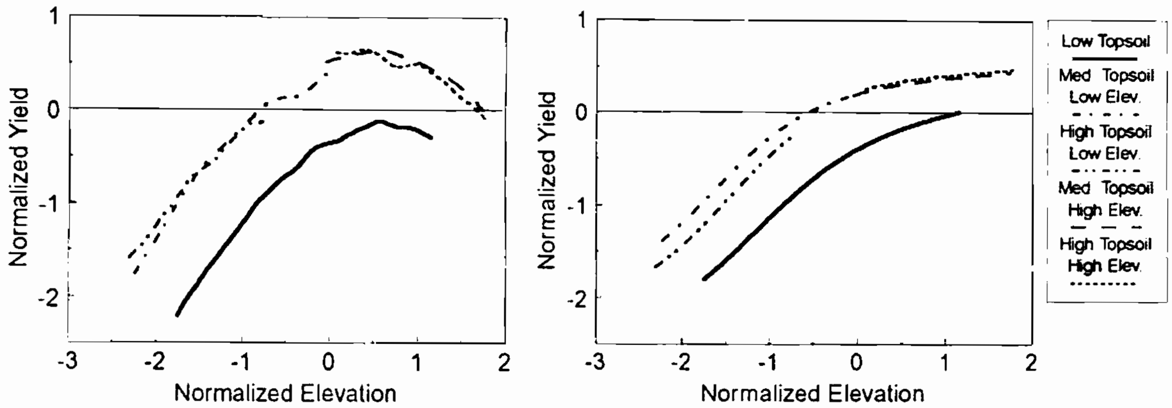

Figure 3. Comparison of PPR-estimated (left) and BPN-estimated (right) soybean yield responses to elevation for Field 2 in 1995.

The PPR response curves appeared to successfully model major yield-limiting factors, as shown by 1995 data from Field 2. For example, higher soybean yields were related to increases in elevation within the field (Figure 3), with the strongest response found in the sub-field areas of lower elevation. This trend was caused by the excess rainfall in the spring of 1995, which caused significant problems with crop stands in the low-lying areas of the field. Yield decreases indicated at the highest elevations were likely caused by the presence of a tree line which reduced yield in that area.

Yield response to topsoil depth was large and positive in the low and medium topsoil areas of the field (Figure 4). The response was negative in the high topsoil areas because the locations of greater topsoil accumulation were generally the same locations where standing water reduced crop stand early in the season. Yield response to higher levels of phosphorus (Figure 4), potassium, and lime were generally negative for this field-year. Most areas of the field were sufficient in these nutrients, and the negative relationship may have been due to mining of nutrients in the more productive areas of the field.

Differences could be observed when comparing the 1995 soybean response curves for Field 2 (Figures 3 and 4) to the 1994 soybean response curves for Field 1 (Figure 5). For example, in low-elevation areas the response of soybean yield to changes in elevation was negative for 1994, but positive for 1995. Since crop growth was water-limited in 1994, the run-on areas at low elevation were at an advantage; while in 1995 similar run-on areas had crop growth limited by excess water. The phosphorus response was also different between the two field-years, with a positive response to higher P observed in the lower landscape areas during 1994. Soil test P levels were lower on Field 1, and significant areas of the field were within the responsive range. The low- elevation portions of the field showed greater response. since crop growth was not limited by low water availability in those areas.

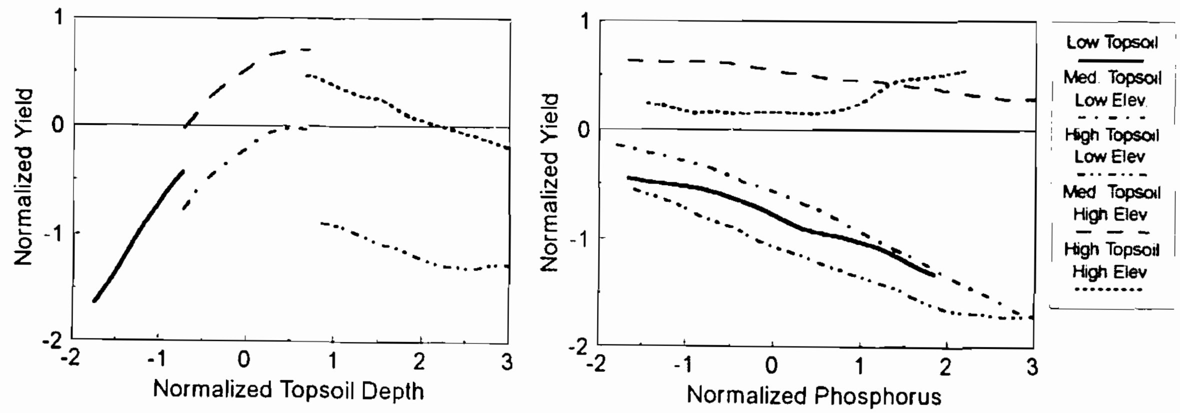

Figure 4. PPR-estimated soybean yield responses to topsoil depth (left) and soil test phosphorus (right) for Field 2 in 1995.

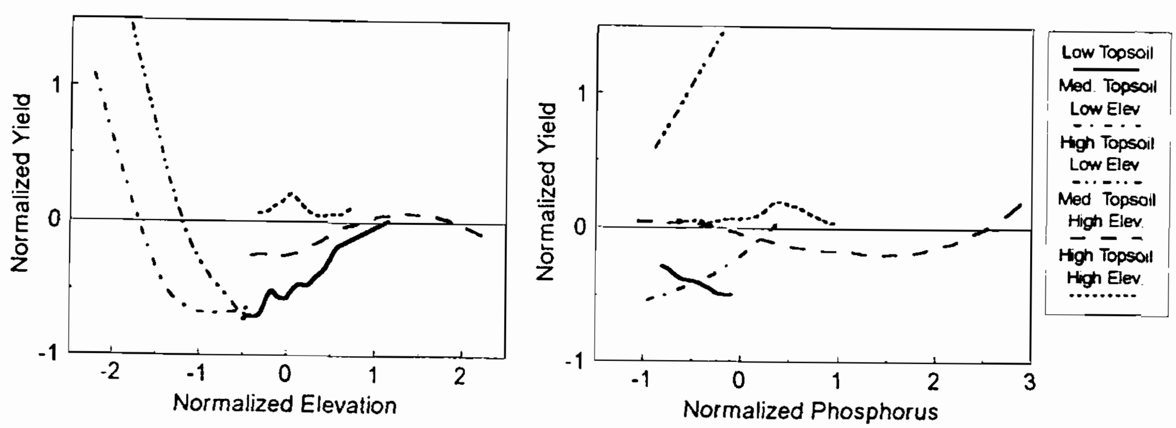

Figure 5. PPR-estimated soybean yield responses to elevation (left) and soil test phosphorus (right) for Field 1 in 1994.

CONCLUSIONS

The process of understanding yield variability was made difficult by the number of inter-related factors that affect yield. Correlation analysis was not particularly useful in understanding yield variability, due to complex nonlinear relationships between yield-limiting factors. Dividing the field into smaller sub-field areas based upon topsoil depth and elevation did not measurably improve the ability of this method to explain yield variability, but other methods of sub-field classification could be investigated in hopes of finding regions that respond homogeneously to input factors. Stepwise multiple linear regression was not able to accurately model yields. As with standard correlation, this was likely due to linearity constraints.

Prediction capabilities were highest for the nonlinear, nonparametric methods. Yield response curves developed with these methods agreed in general with our observations of yield-limiting behavior on these fields. These response curves will be useful for studying the interactions between multiple critical factors and crop yields. The fact that similar predictions and response curves were obtained from two dissimilar methods (PPR and BPN) provides some degree of confidence that the methods are modeling the association between yield patterns and soil properties in a reasonable manner. Further study and refinement of both methods is needed, to optimize their use for yield response investigation.

REFERENCES

Birrell, S.J., K.A. Sudduth and S.C. Borgelt. 1995. Crop yield mapping: comparison of yield monitors and mapping techniques. Comp. Elect. Agric. 14:215-233.

Buchholz, D.D. 1983. Soil Test Interpretations and Recommendations Handbook. Univ. of Missouri, Coll. of Agric., Dept. of Agronomy, Columbia, MO.

Drummond, S.T., K.A. Sudduth, and S.J. Birrell. 1995. Analysis and correlation methods for spatial data. Paper No. 951335. ASAE, St. Joseph, MI

Friedman, J.H. and W. Stuetzle. 1981.Projection pursuit regression. J. Am. Stat. Assoc. 76:817-823.

Hoogenboom, G., H. Lai and D.D. Gresham. 1993. Spatial yield prediction. Paper No. 933350. ASAE, St. Joseph, MI.

Hopfield, J.J and D.W. Tank. 1986. Computing with neural circuits: a model. Science 233:625-632.

Kitchen N.R., K.A. Sudduth and S.T. Drummond. 1996. Mapping sand deposition from 1993 midwest floods with electromagnetic induction measurements. J. Soil Water Cons. 51(4):336-340.

Sudduth, K.A., N.R. Kitchen, D.F. Hughes, and S.T. Drummond. 1995. Electromagnetic induction sensing as an indicator of productivity on claypan soils. *In* Proc. Second Intl. Conf. on Site-Specific Management for Agricultural Systems, pp. 671-681. Am. Soc. Agron., Madison, WI.

PROCEEDINGS OF THE TWENTY-SIXTH

NORTH CENTRAL EXTENSION-INDUSTRY

SOIL FERTILITY CONFERENCE

Published for
The North Central Extension-Industry Soil Fertility Conference
by
Potash & Phosphate Institute
700 - 22nd Avenue, South
Brookings, SD 57006
605-697-7149

November 20-21, 1996

St. Louis Westport Holiday Inn
St. Louis, Missouri

Volume 12

Program Chairman and Editor:

Dr. Peter Scharf
University of Missouri
Dept. of Agronomy
Columbia, MO 65211