Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Computers and Electronics in Agriculture 46 (2005) 309–327
Computers and electronics in agriculture
www.elsevier.com/locate/compag

Identifying potential soybean management zones from multi-year yield data

Dan B. Jaynes*, Tom S. Colvin, Tom C. Kaspar

National Soil Tilth Laboratory, USDA-ARS, 2150 Pammel Dr., Ames, IA 50011, USA

Abstract

One approach for developing potential management zones for a variable-rate precision-agriculture system is to identify areas within a field exhibiting similar yield behavior. In this study, we applied cluster analysis of multi-year soybean (*Glycine max* [L.] Merr.) yield to partition a field into a few groups or clusters with similar temporal yield patterns and investigated the relationships between these yield clusters and the easily measured properties elevation (and the simple terrain attributes derived from elevation) and apparent soil electrical conductivity (EC_a). The analysis was applied to 5 years of soybean yield data collected from 224 plots arranged along eight transects spanning a 16-ha field. The partitioning phase of cluster analysis revealed that the 224 locations were best grouped into five clusters. These clusters were roughly aligned with landscape position and were characterized by the yield response to growing season precipitation above or below the 40-year average. Canonical discriminant functions constructed from the simple terrain attributes and EC_a predicted correct cluster membership for 80% of the plots. While not perfect, the discriminant functions were able to capture the major characteristics of the yield cluster distribution across the field, indicating that these easily measured variables are strongly related to soybean yield. Applying the functions with high-resolution terrain and EC_a attributes, we mapped soybean yield zones within the 16-ha field and an adjacent 16-ha field where multi-year yield data were not available. Cluster analysis of multi-year yield data and easily measured terrain and soil date may be useful in constructing effective management zones within fields and once developed can be applied to similar fields lacking detailed spatial yield data. Published by Elsevier B.V.

Keywords: Yield zones; Clustering; Soybean; Yield variability; Terrain; Soil electrical conductivity

* Corresponding author. Tel.: +1 515 294 8243; fax: +1 515 294 8125.
E-mail address: jaynes@nstl.gov (D.B. Jaynes).

0168-1699/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.compag.2004.11.011

1. Introduction

One obstacle to applying precision agriculture practices to optimize crop production and environmental quality is identifying management zones—areas within the field by which inputs are managed to optimize economic return or environmental impact. Numerous methods have been used to construct potential management zones, but they essentially fall into three types. One approach is to use soil properties, e.g., soil series, water holding capacity, organic matter content, texture, depth to restricting layer, soil fertility test information—to construct potential management zones (Wibawa et al., 1993; Anderson and Bullock, 1998; Van Alphen and Stoorvogel, 2000; Ferguson et al., 2002). This approach assumes that the soil properties that control yield response to inputs are known and measurable. A major limitation of this approach is that extensive soil sampling, often in grid patterns, is required which can be costly and labor intensive.

The second approach is designed to circumvent the cost and time required to collect extensive soil data and instead uses surrogate variables to construct potential management zones. Typically, these surrogate variables include elevation and the simple terrain attributes that can be easily calculated from digital elevation data, such as slope and curvature, as these often account for much of the soil and yield variation observed within fields (Halvorson and Doll, 1991; Afyuni et al., 1993; Brubaker et al., 1993; Timlin et al., 1998; Yang et al., 1998; Kravchenko et al., 2000; Fraisse et al., 2001; Kaspar et al., 2003). Apparent soil electrical conductivity (EC_a) is another surrogate variable often used as it is has been found to be correlated with soil properties that affect yield (Rhoades and Corwin, 1981; Williams and Hoey, 1987; Kachanoski et al., 1988; McBride et al., 1990; Jaynes et al., 1995b) and has been found to be highly correlated with yield (Jaynes et al., 1995a). Apparent electrical conductivity can be measured for fields rapidly and easily using either electromagnetic induction instruments (Jaynes et al., 1993) or direct contacting equipment (Lund et al., 1999). While this approach circumvents the necessity to collect costly soil measurements, it assumes a strong correlation between the surrogate variables and the soil properties that control yield response to inputs.

The third approach makes no assumption regarding the interaction between yield and soil or landscape properties but instead uses the yield data directly to identify areas within a field where crops respond similarly over years (Lark and Stafford, 1997; Stafford et al., 1999; Lark, 2001). This approach makes the assumption that if yield patterns are similar over time then the areas must respond similarly to weather variability and management inputs and may function as effective management zones. Developing management zones from multi-year yield data is an intuitively attractive approach because it relies on direct observations to define yield zones rather than assuming a relationship between yield and soil or surrogate data. Of course, all of these approaches merely identify potential management zones. Further research is required to test whether or not the management zones identified by any approach do in fact function as effective management zones for the application of inputs.

Recently, Jaynes et al. (2003) took the third approach for developing potential management zones by applying unsupervised cluster analysis to multi-year corn yield data. They found that a 16-ha field could be partitioned into a few areas or zones where yield patterns over multiple years were similar. In the interpretation phase of cluster analysis, they found that these zones were largely determined by yield response to growing season precipitation

being either above or below the long-term average. Finally, they used the profiling phase of cluster analysis to determine which, if any, easily measured surrogate variables were useful in predicting the distribution of the potential management zones within the field. They demonstrated a strong relationship between EC_a and simple terrain variables and showed how these could be used to interpolate the distribution of the potential yield zones within areas of the field where yield was not measured.

In this study, we apply cluster analysis to 5 years of soybean yield data measured along multiple transects within the same 16-ha field used in the study by Jaynes et al. (2003). We then use interpretation techniques to investigate the spatial and temporal characteristics of the resulting soybean yield clusters. In the final profiling step of cluster analysis, we use discriminant functions based on appropriate terrain and EC_a data to give insight into the potential processes and properties that control soybean yield and show how discriminant functions can be used to map soybean yield zones in areas where long-term yield data does not exist.

2. Materials and methods

The study was conducted on a 16-ha field in central Iowa (42°05′N, 93°46′W; Fig. 1) first described by Steinwand and Fenton (1995) and most recently discussed in Kaspar et al. (2003) and Jaynes et al. (2003). Soils in this field were formed in young glacial till of the Des Moines lobe and are in the Clarion (fine-loamy, mixed, and mesic Typic Hapludolls) – Nicollet (fine-loamy, mixed, mesic, and Aquic Hapludolls) – Webster (fine-loamy, mixed, mesic, and Typic Haplaquolls) association (Steinwand and Fenton, 1995). The field has a gently rolling topography typical of the Des Moines lobe. The field had been in a 2-year rotation of corn and soybean since 1957 with field management typical for central Iowa (Karlen and Colvin, 1992; Colvin et al., 1997).

We measured corn and soybean yields for 11 consecutive years starting in 1989. Grain yield was measured along eight east–west transects spaced 48.8 m apart with a combine

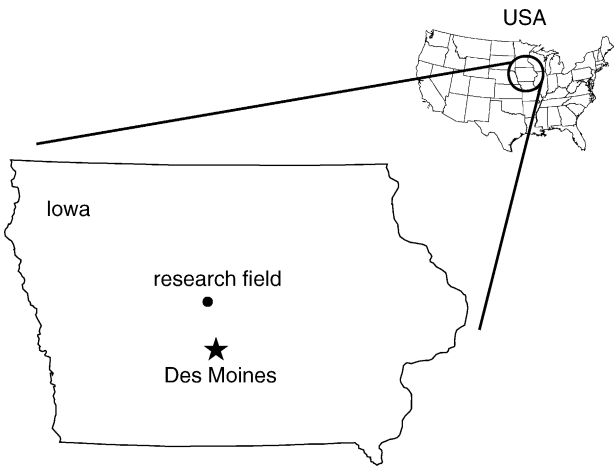

Fig. 1. Location of research field in central Iowa, USA.

modified to support a weigh hopper mounted inside the grain-storage tank (Colvin, 1990). A width of 2.28 m (three rows) for corn and 3.81 m for soybean was harvested along each transect. Crops were harvested by driving the combine for a measured distance and then stopping to weigh the accumulated grain and measure its moisture content. While the position and length of each transect were consistent from year-to-year, the exact length of individual plots within the transects varied in some years, but averaged slightly over 12 m. Yield data were linearly interpolated to correspond to plots of uniform 12.1-m length, resulting in 28 plots per transect. Thus, 224 (8×28) yield values on a grid were obtained for each year (Fig. 2a). Corn yields were adjusted to a moisture content of 155 g/kg and soybean yields were adjusted to moisture content of 130 g/kg. Starting in 2000, the field was combined with the 16-ha field to the north and the entire 32-ha field planted to a corn and soybean rotation with rows running N–S instead of E–W. Soybean yield in 2000 was measured with the producer's combine fitted with a yield monitor and GPS system.

2.1. Field attributes

Apparent soil electrical conductivity and elevation were measured across the combined 32-ha field after soybean planting early in June 2000 when the soil profile was near field capacity. Elevation and position measurements were made with a kinematic DGPS receiver (Ashtech Z Surveyor, Magellan Corp., Santa Clara, CA[1]) mounted on an all-terrain vehicle (ATV). Apparent soil electrical conductivity was measured inductively using an EM-38 electrical conductivity induction meter (Geonics Ltd., Mississauga, Ont., Canada). The EM-38 was pulled behind the ATV attached to a fiberglass boom (Jaynes et al., 1993, 1995a). The meter was attached to the boom in the vertical dipole position and maintained 8 cm above the ground surface. Apparent electrical conductivity was measured and recorded with the position and elevation data. Readings were logged every 1 s as the ATV moved across the field at approximately 4 m s^{-1} giving measurements about every 4 m. North-to-south transects were driven approximately 8.7 m apart across the field. Elevation data was supplemented by collecting additional elevation data by driving along ridges and swales to minimize interpolation errors in the subsequent terrain model. A base-station GPS receiver, located at a benchmark on the eastern edge of the field, was used to differentially correct the roving GPS receiver. Position measurements were reliably within ± 0.03 m horizontally and ± 0.06 m vertically for this equipment.

Position data were referenced to a Universal Transverse Mercator (UTM) projection (Zone 15, North American Datum 1983). Elevation values were estimated in height above the ellipsoid (m). The elevation data were used to generate an 8 m $\times$ 8 m digital elevation model (DEM) using Surfer 7.0 (Golden Software Inc., Golden, CO) gridding software and a linear, isotropic variogram model. The grid was then smoothed using a cubic spline procedure in Surfer 7.0 creating a 2 m $\times$ 2 m DEM. The smoothing procedure was used to reduce the roughness of the DEM caused by small DGPS measurement errors in elevation. The primary terrain attributes: EL, elevation (m); SL, slope (the rate of maximum change in elevation

[1] Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

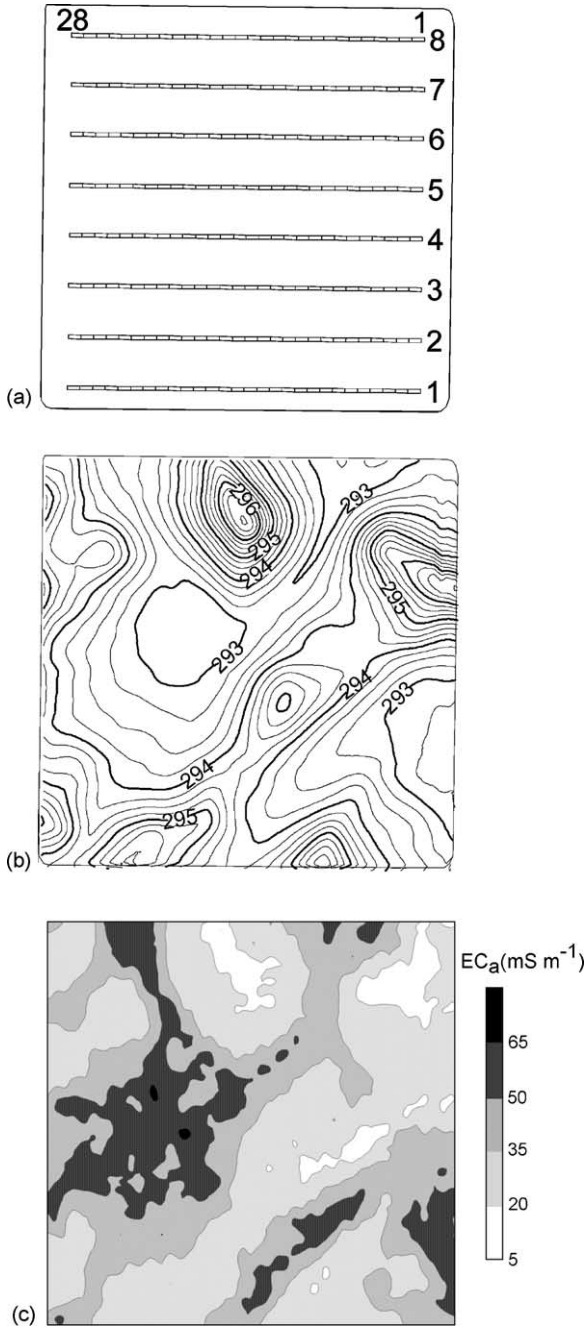

Fig. 2. Schematic of 16-ha field showing (a) location and numbering scheme for yield plots and transects where yield measurements were taken within the field, (b) elevation contours in m above mean sea level for the 16-ha field, and (c) EC_a interpolated from electrical conductivity induction measurements.

to surrounding grid cells, E); PL, plan curvature (curvature of the surface perpendicular to the direction of slope, ($100\,m^{-1}$); PL < 0 for curvatures that are concave upwards); PR, profile curvature (curvature of the surface in the direction of the slope, ($100\,m^{-1}$), PR < 0 for curvatures concave upwards); AS, aspect (absolute deviation of the slope direction from south, E) were then calculated for each 2-m grid cell of the DEM using the Arc/Info GIS software CURVATURE command (Arc/Info, 1998; Environmental Systems Research Institute, Redlands, CA). To quantify the effect the closed surface depressions could have on yield, a depression depth (DD) attribute was calculated. The DEM fill script associated with the Spatial Analyst extension of ARCView was used to numerically fill in the internal depressions within the DEM. This new surface was then subtracted from the original DEM to give the depth of depressions. Values for DD are typically equal to 0 over most of a field but greater than 0 in areas with closed depressions. A 2-m grid cell coverage of EC_a for the field was created from the survey data using the Arc/Info GIS TOPOGRID command and an isotropic, Gaussian variogram model.

The 224 transect plots were digitized as polygons and overlaid on the EC_a and terrain coverages. The area within each of the 224 transect plot polygons was converted to a raster format with a 0.25-m^2 resolution to better align with the transect plot borders. The values of the field attributes for each 0.25-m^2 yield-polygon raster were taken from the underlying 2-m grid cells of the appropriate coverage. The mean value for each attribute within each of the 224 transect plots was calculated by arithmetic averaging of the values for all 0.25-m^2 rasters that fell within the specific transect plot polygon.

2.2. Cluster analysis

Cluster analysis is a three-part process comprised of partitioning, interpretation, and profiling. In the partitioning step, a clustering algorithm is used to divide the members of a population into one of several clusters or groups such that the differences among groups are minimized while the differences between different groups are maximized. For yield data, partitioning groups the transect plots into clusters having similar yield patterns for the 5-year period, while minimizing the similarity in yield patterns between clusters. After partitioning, an interpretation step is used to examine the yield characteristics that led to the formation of the clusters in order to describe and label the nature of the different clusters. For yield clustering, this includes characterizing the spatial and temporal nature of the cluster. The interpretation step can provide valuable insights into the yield characteristics of each cluster. Profiling is the final step of clustering analysis and is used to relate the characteristics of each cluster in terms of auxiliary data such as soil properties not used in the partitioning phase. Profiling focuses on determining not what directly determined the clusters but on the secondary characteristics of the clusters after they are identified. Here, we used profiling to investigate the relationships between the yield clusters and the easily measured surrogate variables of terrain and EC_a. By using discriminant analysis for the profiling phase, significant relationships can be used to predict the occurrence of yield clusters in areas of the field where yield data is lacking. The discriminant functions can also be used to predict the occurrence of yield clusters in other fields that lack multi-year yield data but have similar soil and landscape characteristics.

Nonhierarchical cluster analysis was performed for soybean yield using data from the 224 transect plots and PROC FASTCLUS (SAS, 2000). Nonhierarchical cluster analysis forms independent groups without assuming a hierarchy or tree structure of interconnections. First, the 5 years of soybean yield data were standardized by subtracting the yield median and dividing by the interquartile range. Global values for median and interquartile range were computed using data from all 224 transect plots and all years. The transect plots were partitioned into clusters using a *K*-means algorithm (Hair et al., 1987) to minimize the sum-of-squares of the standardized yields for cluster members. Cluster analysis was used repeatedly to group the plots into 2, 3, 4, 5, 6, 7, and 8 clusters to determine the optimum number of clusters to use. We ended with eight clusters as we assumed eight to be the maximum number of zones to be of practical use for management decisions. The cluster algorithm was run for 20 iterations or until the change between iterations in all cluster means was 0. A pseudo *F* statistic (Milligan and Cooper, 1985) computed as

$$\text{pseudo } F = \left(\frac{R^2}{c-1}\right)\left(\frac{(n-c)}{1-R^2}\right)$$

where R^2 is for overall prediction of yield by cluster, c the number of clusters, and n is number of observations was used as an indicator for the optimum number of clusters.

For the interpretation step of cluster analysis, the spatial structure of the resulting clusters was quantified using Moran's I statistic (Moran, 1950; Upton and Fingleton, 1985). Moran's I is similar in concept to correlation and ranges from −1 to 1. A Moran's I near −1 indicates that members of different clusters are evenly interspersed across the field like the colored squares of a checkerboard. A Moran's I = 0 indicates a completely random distribution of the clusters and a value near 1 indicates that members of a cluster are grouped closely together in space. Moran's I was calculated using the Excel 97/2000 Visual Basic routine written by Sawada (1999). The spatial pattern of the yield clusters was also examined qualitatively by overlaying the yield clusters on the elevation contour map of the field. Interpretation of the temporal pattern of the clusters was determined by one-way analysis of variance (PROC ANOVA; SAS, 2000) of the soybean yields for each year with cluster as the main effect. Where the ANOVA *F*-test was significant, differences in the mean soybean yields for each cluster were tested using Duncan's multiple range test ($P = 0.05$).

2.3. Multiple discriminant analysis

We used discriminant analysis for the profiling step to quantify the relationship between the yield clusters and the terrain attributes and EC_a. Discriminant analysis is appropriate to use when the dependent variable (clusters) is categorical and the independent variables are continuous (Hair et al., 1987). Given a set of yield clusters, discriminant analysis develops functions of the field attributes that most effectively discriminate between the yield clusters. Discriminant analysis was also used to develop canonical composites of the field attributes that best discriminated between the yield clusters. A maximum of $n - 1$ canonical composites can be computed to distinguish between n clusters. The canonical composites and the correlation or loading of each field attribute to the composites were used to examine the combination of field attributes that influenced yield patterns (Hair et al., 1987).

To start the discriminant process, a forward step-wise discriminant analysis (PROC STEPDISC; SAS, 2000), similar to forward step-wise regression, was performed to determine which readily measured field attributes contributed towards classifying the transect plots into yield clusters. Values for EC_a, EL, SL, PR, PL, AS, and DD averaged for each yield plot were included in the analysis. In the first step, the field attribute that contributed most to discriminating between the clusters was brought into the discriminant model. In the second step, the attribute that contributed most to the discriminating power of the model, which was not already in the model, was entered into the model if it exceeded the preset significance level for entry. Before each new attribute was entered into the model, the method tested the attributes already in the model. If the attribute that contributed the least to the model failed to satisfy a preset significance level for remaining in the model, it was removed. The step-wise process continued until all attributes in the model meet the criterion to stay and none of the remaining attributes meet the criterion to enter the model. A moderate significance level ($P = 0.15$) was used for attributes both entering and leaving the model because this level was mid-range of the levels found to perform best in the forward selection methods (Costanza and Afifi, 1979).

Field attributes found to be significant in the step-wise discriminant analysis were used to develop functions to discriminate between clusters. The data for the 224 plots were randomly divided into subsets such that each subset had a similar number of members from each cluster. The first data subset was used as a calibration dataset to develop a set of functions based on the field attributes that best discriminated between the yield clusters (PROC DISCRIM; SAS, 2000). The accuracy of these functions was tested by predicting cluster membership of the yield plots in the second, validation data subset and comparing with known membership. The calibration–validation process was repeated by reversing the roles of the two data subsets (double cross-validation), giving two estimates for the accuracy of the discriminant functions in predicting proper cluster membership. Calibrating and validating the discriminating functions on two different subsets of the data eliminates the upward bias of testing the functions on the same data set for which they were calibrated (Hair et al., 1987).

After evaluating the accuracy of the discriminant functions, a set of $n - 1$ canonical discriminant functions was computed using the combined data from all 224 transect plots. The canonical composites and the loadings of each field attribute to the composites were used to identify those combinations of field attributes that are important for explaining the yield cluster pattern within the field. Finally, the discriminant functions were used in combination with the 2-m resolution field attribute data to predict the spatial distribution of yield zones across the southern and northern 16-ha fields where detailed yield data were lacking.

3. Results

3.1. Field measurements

The elevation contour and interpolated EC_a map for the southern 16-ha are shown in Fig. 2. Total relief within the southern half of the field was about 4.5 m. The field topography

Table 1
Univariate statistics for soybean grain yield within 224 transect plots, and growing season precipitation (May–August) by year and 5 years combined

Statistic	1990	1992	1994	1996	1998	5 years combined
Yield (Mg ha^{-1})						
Mean	3.16	3.08	3.12	2.95	2.75	3.01
Standard deviation	0.96	0.36	0.56	0.95	0.93	0.81
Maximum	4.28	3.71	4.11	4.11	3.82	4.28
Minimum	0.00	1.67	1.15	0.00	0.00	0.00
Skew	−2.20	−1.25	−1.63	−1.97	−1.76	−2.18
Kurtosis	4.41	1.78	2.69	3.10	2.62	5.10
Median	3.48	3.14	3.29	3.29	3.10	3.23
Interquartile range	0.67	0.39	0.50	0.61	0.89	0.59
Precipitation (mm)[a]						
Sum	732	357	358	555	528	506

[a] 40-year average precipitation is 442 mm.

within this area was dominated by a closed depression or pothole in the west-central region (maximum DD = 0.38 m) and by hills in the north-central and northeast areas. Occurrence of potholes is typical in this geologically young landscape that is characterized by a poorly defined surface drainage system (Andrews and Dideriksen, 1981). A low ridge spanned the eastern to south-central portion of the field. Apparent electrical conductivity varied from 9.9 to 67.9 mS m^{-1}, which we have found to be typical for springtime EC_a surveys of soils in this area. Apparent electrical conductivity was higher in the lower portions of the landscape where more poorly drained, finer-textured soils high in soil organic matter predominate. Lower EC_a values were measured near hill and ridge tops where well-drained, coarser-textured, eroded soils are common in this landscape. These trends reflect the expected general relationship of higher EC_a values in areas with higher soil–clay contents, which are also the areas with higher water contents when the field is near field capacity.

Soybean yield from the transects and growing season precipitation (May–August) are given in Table 1. Wetter than average weather in 1990, 1996, and 1998 caused complete loss of soybean yield in some transect plots due to temporary flooding and reduced the field-average yield in 1998. Zero or very low yields in the wetter years were centered on the pothole in the west central region of the field (Fig. 2) where runoff occasionally ponded for several days. The maximum average yield occurred in 1990, which was also the year when the highest yield for any transect plot was measured. Yields were markedly more uniform in the drier years of 1992 and 1994, which was also observed for corn yields in this field (Jaynes et al., 2002; Kaspar et al., 2003). Yield distributions in every year were negatively skewed with the median exceeding the mean, and failed the Kolmogorov–Smirnov test for normality.

3.2. Partitioning

Partitioning the 5 years of soybean yield data into two to eight clusters gave five as the optimum number of clusters as determined by local maximum values for the pseudo F

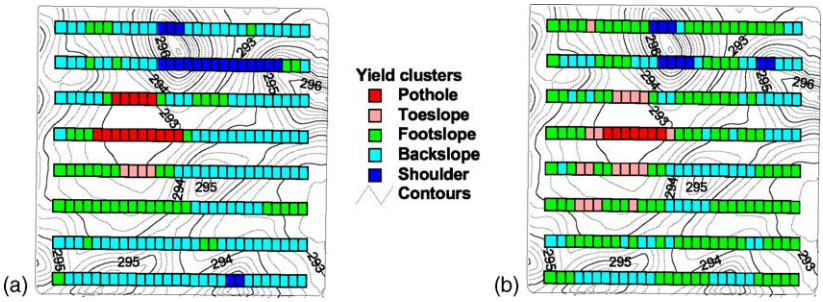

Fig. 3. Yield cluster classification for the 224 transect plots overlaid on the elevation contours for (a) soybean and (b) corn. Transect plots shown 3× actual width for better visibility. Corn yield clusters from Jaynes et al. (2003).

statistic (results not shown). Dividing the field into more than five clusters tended to create clusters with only one or two members—a result not desireable for clustering analysis (Hair et al., 1987). Yield plot membership within the five clusters ranged from 4 within one cluster to 135, or more than half of the total number of transect plots, in another.

3.3. Interpretation

The spatial distribution of the five clusters was not random across the field, but appeared to form contiguous areas instead (Fig. 3a). Moran's I for the soybean cluster distribution was 0.74 ($P < 0.001$) confirming that the clusters tended to group together within the field. It is important to note that this spatial correspondence was not a direct result of the clustering algorithm, which partitioned the plots based on yield patterns over the 5 years but used no spatial information. Rather, the resulting spatial structure of the clusters reflected spatial correlation of some underlying property or process that affected yield. Had the yield clusters formed a more random spatial pattern, variable rate application of inputs would not be a viable management alternative for this field.

The distribution of soybean yield clusters within the field loosely followed landscape position (Fig. 3a). One cluster had members that were located entirely within the pothole area of the field and for convenience will be referred to as the Pothole cluster. Adjacent to the pothole, but slightly higher in the landscape was a small extent of a second cluster we will call the Toeslope cluster. A third cluster contained yield plots at lower elevations that were not located within closed depressions and will be called the Footslope cluster. Members of a fourth, Shoulder cluster, occupied scattered Shoulder and hill top locations within the field. The fifth or Backslope cluster contained transect plots that were generally higher that the Footslope plots and lower than the Shoulder plots. That the yield clusters would be roughly congruent with landscape positions is not surprising for this field. Developed in uniform parent material, most differences in soil across the field were caused by soil forming factors driven by topography. After farming commenced on this field, erosion would have been greatly accelerated and this too would have been controlled by topography, with sediment moving from the higher, steeper locations and depositing in the lower and depressional areas of the field (Pennock and de Jong, 1987).

Table 2
Deviation of growing season rainfall above or below 40-year average for each year, and means of soybean yield by year and cluster

Year	Rainfall	Mean yield (Mg ha^{-1})				
		Pothole	Toeslope	Footslope	Backslope	Shoulder
1990	Above	0.26 e[a]	1.67 d	3.19 b	3.62 a	2.48 c
1992	Below	3.14 b	3.56 a	3.25 b	3.10 b	2.29 c
1994	Below	3.55 a	3.55 a	3.21 b	3.22 b	1.69 c
1996	Above	0.28 e	0.70 d	3.00 b	3.44 a	1.90 c
1998	Above	0.00 d	2.56 b	2.35 bc	3.30 a	2.12 c

Analysis of variance results were significant ($P < 0.001$) for each year.
[a] Means within a row followed by same letter are not significantly different at $P = 0.05$ using Duncan's multiple range test.

Differences in soybean yield for each cluster and year were tested by one-way analysis of variance (Table 2). In every year, there was a significant difference in the soybean yield among the clusters ($P < 0.001$). Testing of means revealed a simple temporal pattern in the soybean yields averaged by cluster. In years with above average growing season precipitation (1990, 1996, and 1998), the Backslope cluster had the highest mean soybean yield and the Pothole cluster had the lowest mean yield of the five clusters. In 1990 and 1996, the pattern in mean yield for the other three clusters was Footslope > Shoulder > Toeslope. In 1998, the yield pattern for the three clusters was reversed with Toeslope > Shoulder and neither the Toeslope nor Shoulder cluster significantly different than the Footslope cluster. In the 2 years with below average growing season precipitation (1992 and 1994), the pattern for the mean yields within the clusters was markedly different (Table 2). In these years, the Toeslope cluster had greater mean yield than the Footslope, Backslope, and Shoulder clusters and the mean yield for the Pothole cluster equaled or exceeded these three clusters. The Shoulder cluster had the lowest mean yield of all clusters in years with below average growing season precipitation.

That the cluster yields exhibited different patterns depending on growing season precipitation being above or below the 40-year average, implies that the partitioning of the yield plots was strongly determined by yield response to soil moisture. Yield clusters lower in the landscape (Pothole and Toeslope) had relatively lower soybean yields in wetter years than in drier years, most likely because of excess soil moisture as well as observed short-term flooding of the pothole. Toeslope yields were nearly as poor as Pothole yields in 1996, but significantly better in 1990, and much better in 1998, indicating that Toeslope clusters were not as affected by excess soil moisture or flooding. In the years that growing season precipitation exceeded the average, 1998 had the smallest excess, which may explain why Toeslope yields were better that year. But growing season precipitation was 180 mm greater in 1990 than 1996 while the Toeslope yields were worse in 1996 than 1990. Thus, deviation from average growing season precipitation was a useful indicator of yield pattern, but knowing that the weather was wetter than average was not sufficient for determining average yield of the Toeslope cluster relative to the other clusters.

The Footslope cluster had nearly the same average yield as the Backslope cluster in the drier years. But, in wetter years, average yield of the Footslope plots were lower than average

yield of Backslope plots perhaps due to excessive soil moisture reducing soil aeration in the lower areas of the landscape (Logsdon et al., 1999) caused by runoff from higher elevations and interflow. Transect plots in the Shoulder cluster consistently had average yields lower than the field average (Table 1) regardless of the amount of growing season precipitation. These plots occupied landscape positions that typically are the most eroded and depleted of soil organic matter—conditions that have been found to reduce yield (Olson and Nizeyimana, 1988; Thompson et al., 1991).

Jaynes et al. (2003) used the same partitioning method and found that corn yields within this field also formed five clusters that roughly aligned with the same landscape positions as found for soybean (Fig. 3b). In addition, the temporal patterns within the corn clusters also reflected the deviation from average of growing season precipitation. Although similar, the distribution of soybean and corn yield clusters within the field exhibited some differences. The number and extent of Pothole yield clusters for soybean expanded slightly and the Toeslope clusters reduced in number in comparison with the equivalent corn clusters. Also, some transect plots nonadjacent to the pothole were assigned to the Toeslope corn yield cluster during partitioning, which did not occur for soybean. This shift in the extent of the clusters impacted by above average precipitation agrees with other research that shows that soybean is less susceptible to yield loss from wet soil conditions than corn (Evans and Fausey, 1999).

The number of members in the Backslope and Shoulder clusters was greater and the number of members in the Footslope cluster was lower for soybean compared with corn. For soybean, more transect plots fell within the Backslope cluster (60%) than any other cluster. While for corn, the Footslope cluster contained the most transect plots (54%). Because yields in the Footslope plots were relatively lower in wetter than average years, this again implies that soybean yields were less affected by wet weather than corn or that the wet years when soybean was grown were not as wet or had different patterns of precipitation than wet years when corn was grown.

Another marked difference in the soybean cluster distributions occurred for Plots 6–12 along Transect 7 in the NE quarter of the field (Fig. 3). These plots partitioned into the Shoulder cluster for soybean, even though they traversed a swale in the landscape. This is in marked contrast to the other members of this cluster, which occupied hilltop and Shoulder locations. Also, partitioning of corn yields placed these transect plots into Backslope and Footslope clusters as would be expected from their position in the landscape. Clearly, factors other than landscape position, such as weeds or disease, affected yield in this area for soybean but not corn. For example, soybean cyst nematodes are known to infest this field and could have lowered soybean yields in some areas of the field while not affecting corn yields even with the use of resistant cultivars.

3.4. Profiling

That the soybean clusters generally followed landscape position suggested that terrain attributes might serve as effective surrogate measures for yield behavior. Discriminant analysis was conducted to determine if the easily measured surrogate variables, EC_a and the terrain attributes, differed by yield cluster and were adequate for predicting the membership of the plots within the different yield clusters. Step-wise discriminant analysis indicated that

Table 3
Number of transect plots in each test data set by soybean yield cluster, and the number of these transect plots predicted to fall into the five yield clusters using discriminant functions derived from EC_a and terrain attributes

Known yield cluster	Test data set #	Plots in zone	Predicted plots in yield cluster				
			Backslope	Pothole	Toeslope	Shoulder	Footslope
Backslope	1	67	60	0	0	2	5
	2	68	57	0	0	1	10
Pothole	1	8	0	5	3	0	0
	2	7	0	6	0	0	1
Toeslope	1	2	0	0	1	0	1
	2	2	0	0	1	0	1
Shoulder	1	9	3	0	0	6	0
	2	10	2	0	0	7	1
Footslope	1	26	7	0	0	1	18
	2	25	6	0	0	0	19

EC_a and the terrain attributes, SL, PL, AS, and DD were significant in determining cluster membership (analysis not shown). This was somewhat different than the results of Jaynes et al. (2003) for corn yield clusters where the terrain attributes EL and PR were also found to be significant. They did not include the terrain attribute DD in their analysis. As a result of the step-wise analysis, canonical discriminant analysis was conducted for soybean clusters using only EC_a, SL, PL, AS, and DD as explanatory variables.

Using the double cross-validation approach, two independent estimates of the accuracy of the canonical discriminant functions were made. Overall, the discriminant functions developed from EC_a and terrain attributes were able to correctly predict cluster membership for 80% of the transect plots for both data subsets (Table 3). Half of the Toeslope members were classified incorrectly by the canonical functions using either half of the data, but this yield cluster only contained a total of four members. Membership in the other yield clusters was predicted about equally well, ranging between 62 and 89% accurate. The most common misclassification for the Backslope plots was into the Footslope cluster and most common misclassification for the Footslope plots was into the Backslope cluster. These were the two largest clusters and were adjacent on the landscape and many of the misclassifications occurred at the transition between these two clusters along the transects. Misclassifications of Shoulder plots occurred along Transect 7, Plots 7–11, and the area where soybean clustering differed most from corn clustering. Including EC_a with the terrain attributes did not account for the soybean yield behavior in this area of the field further implying that soybean yields were being affected by either some nonedaphic condition or a soil property not correlated well with EC_a.

The 80% prediction accuracy appears high, but is best evaluated against the maximum chance and proportional chance criteria (Hair et al., 1987) to gage the utility of the multiple discriminant functions. The maximum-chance criterion is equivalent to the classification percentage when all the yield plots are classified into the single cluster with the greatest number of members (Backslope cluster). The maximum-chance criterion was 60 and 61% for the two subsets, respectively. The proportional chance criterion was computed by sum-

Table 4
Results of canonical discriminant analysis of the field attribute data for identifying soybean yield clusters showing eigenvalues, proportion of partitioning accounted for by each composite, and the loadings of each field attribute on the composite

	Composite 1	Composite 2	Composite 3	Composite 4
Eigenvalue	5.57	0.81	0.45	0.00
Proportion	0.81	0.13	0.07	0.00
Loadings				
EC_a	0.445	−.541	0.663	−0.164
SL	−0.356	0.834	−0.176	0.286
PL	−0.071	0.703	0.164	−0.182
AS	−0.083	−.246	0.346	0.870
DD	0.995	0.025	−0.074	0.046

ming the squares of the fractional membership in each cluster and was 42 and 43% for the two subsets, respectively. Thus, the classification success percentages for the two sets of discriminate functions exceeded both chance criteria by ≥19%, indicating that the field attributes were useful for classifying yield plots into the correct clusters.

Thus, the discriminant analysis indicated that there was a strong relationship between the terrain attributes and EC_a and the soybean yield clusters, but that these field attributes were not sufficient by themselves to completely account for all of the yield clustering. This failure was the result of the yield patterns not being determined by terrain and EC_a alone. Most likely soil properties not correlated with terrain or EC_a and factors such as disease and weed distributions, or current or past management practices also affected yield.

Canonical discriminant functions were calculated after combining data from all 224 transect plots. Of the four possible composite functions, only the first two were important for identification of cluster membership (Table 4). The first composite accounted for 81% of the discrimination between clusters and was most heavily loaded by DD and secondly by EC_a. This composite was very effective in distinguishing the Pothole and Toeslope clusters from each other and less effective in distinguishing the Footslope from Backslope cluster (Fig. 4). The second canonical composite accounted for 13% of the overall discrimination and was most heavily loaded by SL, PL, and EC_a. This composite was most effective in discriminating plots in the Shoulder cluster from plots in the Backslope cluster. The last two composites accounted for 7% of the overall discrimination and with eigenvalues $\ll 1$, were not very important for identifying cluster membership.

While the field attributes investigated here were unable to completely predict membership, the 80% success rate indicates that spatio-temporal soybean yield patterns within this field are strongly related to processes or properties that are reflected by terrain properties and soil EC_a. These field attributes can also be used as a first approximation of the distribution of yield clusters across the field by using the discriminant functions developed above. Applying these discriminant functions to the 2-m resolution data for terrain attributes and EC_a collected for the combined northern and southern 16-ha fields, yield clusters can be estimated for areas where yield data were not collected (Fig. 5).

Comparing these predicted yield zones with the partitioning results for the transect plots again illustrates the ability of the field attributes to predict spatio-temporal yield behavior. As expected, the soybean yield zones constructed from the discriminant functions and 2-m

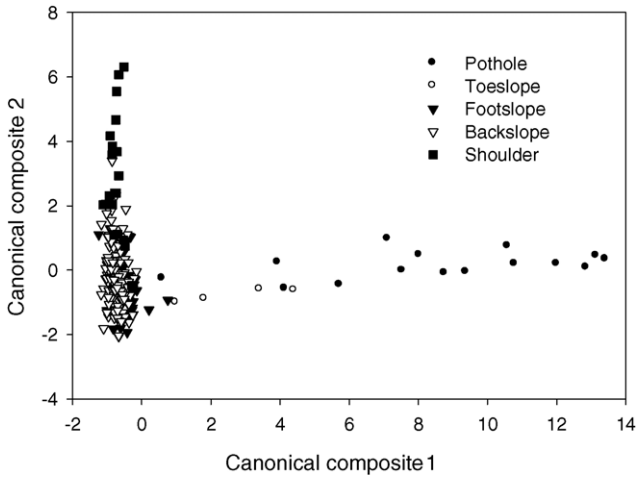

Fig. 4. Separation of the 224 transect plots as determined by the first two canonical discrimant functions constructed from terrain attributes and EC_a. Plots are identified by cluster membership as determined by partitioning.

resolution field attributes did not correctly identify the aberrant lower yields found in Plots 6–12 along Transect 7. Also, transitions from one cluster to another were not always predicted accurately. Overall, however, the yield zone map faithfully captured the yield partitioning of the multi-year soybean data and covered the areas between the transects within the southern field where yield data were not available.

The process also predicted the distribution of soybean yield zones across the northern field where no yield data was collected prior to 2000. No Pothole or Toeslope yield zones were predicted for this field, but the other three zones were well represented. To test the utility of these predicted zones, mean soybean yields for 2000 were computed for the different zones in the southern and northern fields by overlaying the farmer's yield monitor data on the yield zone map in Fig. 5. Growing season precipitation in 2000 was below the 40-year average so we would expect a yield pattern for the yield zones to be similar to patterns in 1992 and 1994. Table 5 shows that the mean soybean yield follows this pattern fairly well. In the southern field, the Pothole and Toeslope yield zones had the highest average yields and the Shoulder yield zone had the lowest average yield. As in the other years with

Table 5
Yield mean and standard error (S.E.) by predicted yield zone for soybean harvested in 2000 from the southern and northern 16-ha fields

Yield zone ($Mg\,ha^{-1}$)	Southern field		Northern field	
	Mean	S.E. (±)	Mean	S.E. (±)
Pothole	3.38	0.24	–[a]	–
Toeslope	3.20	0.22	–	–
Footslope	3.06	0.55	3.04	0.48
Backslope	2.94	0.63	2.84	0.33
Shoulder	2.62	0.86	2.43	0.40

[a] Yield zone not present within field.

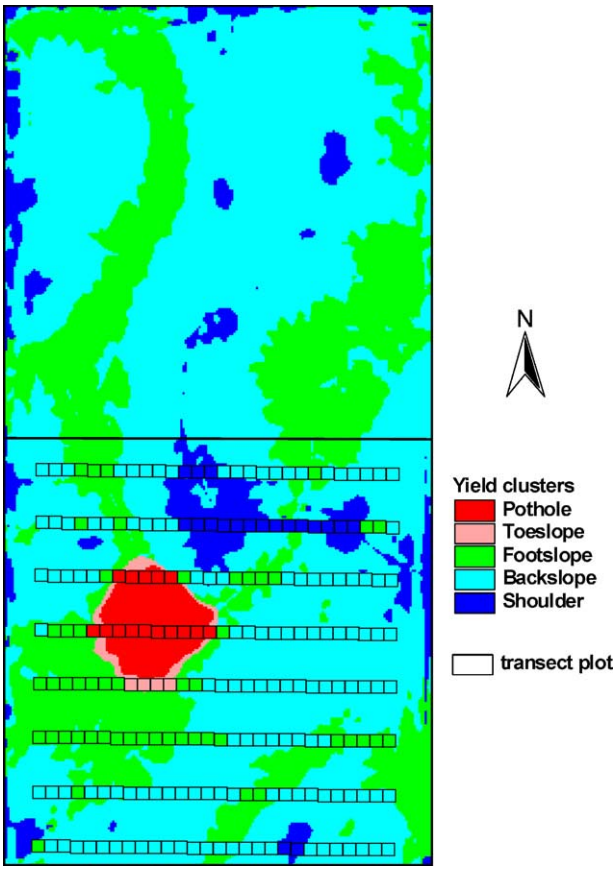

Fig. 5. Yield clusters for the 224 transect plots overlaid on the predicted yield zones for the southern and northern 16-ha fields. Yield zones were determined from the canonical discriminant functions and high resolution terrain attributes and EC_a. Transect plots shown 3× actual size for better visibility.

below average growing season precipitation, the Footslope and Backslope yield zones had similar average yields that lay between these extremes. The North field had a similar pattern although no Pothole or Toeslope yield zones were present. The Shoulder yield zone had by far the lowest average yield and the Footslope and Backslope yield zones had average yields similar to each other and to the average yields in their counterpart zones in the southern field. Thus, the discriminant functions may be valuable in developing soybean yield zones in similar fields where multi-year spatial yield data are not available.

4. Conclusions

Cluster analysis of multi-year soybean yield data partitioned the data into five clusters that roughly aligned with landscape position within a 16-ha field. The clusters were characterized

by the yield response to years having above and below average growing season precipitation and by their rough alignment with landscape position. In growing seasons with greater than long-term average precipitation, yields in the Pothole and Toeslope clusters (i.e., clusters located lower in the landscape) were lower than the field mean. Whereas in years with lower than average precipitation, yields in these clusters were equivalent or better than yields in the Backslope cluster and greater than the field average. Cluster analysis identified a Shoulder zone where yields were always below the field average regardless of precipitation. Soybean clusters were similar to yield clusters found independently for corn in an earlier study (Jaynes et al., 2003). This similarity reflects the role of landscape position in crop growth and yield in rain-fed fields and reflects the general importance of soil moisture, both insufficient and excess, in determining crop yields.

The terrain attributes SL, PL, AS, and DD and EC_a were effective in correctly identifying yield cluster membership for 80% of the 224 transect plots. Misidentification of the other 20% was due to unknown terrain or edaphic properties, pest or disease distributions, or management practices, and illustrates the danger of determining yield zones exclusively from soil or related data as our understanding of all the factors that control spatio-temporal yield distributions is imperfect. Nevertheless, 80% accuracy means that these easily measured field attributes are strongly related to soybean yield and can be used, as a first approximation, to map the distribution of yield zones in similar fields where multi-year spatial yield data are not available.

At this point, we have developed soybean yield clusters using cluster analysis applied to multi-year yield data collected along transects. We have shown that elevation and EC_a, which are easily measured and often effective as surrogate variables for soil properties and processes, are highly related to the yield clusters. Moreover, these variables can be used as disciminant functions to predict spatial distribution of soybean yield clusters in areas with similar landscape and soil characteristics, but where yield data are not available. Yield zones represent areas of the field where soybean plants respond similarly to weather and uniform management across years as indicated by yield and respresent sensible management zones for varying agricultural inputs. However, these zones must be thought of as potential management zones only. A necessary next step is to determine if there are unique relationships between the yield response in these zones and the management of inputs. For example, if the economically optimum fertilizer rate differed by yield zone then these zones would form effective units for fertilizer management. If this next step proves successful, then yield zones could be used effectively as management zones in a variable rate, precision farming program.

References

Afyuni, M.M., Cassel, D.K., Robarge, W.P., 1993. Effect of landscape position on soil water and corn silage yield. Soil Sci. Soc. Am. J. 57, 1573–1580.

Anderson, L.L., Bullock, D.G., 1998. Variable rate fertilizer application for corn and soybean. J. Plant Nutr. 21, 1355–1361.

Andrews, W.F., Dideriksen, R.O., 1981. Soil Survey of Boone County, Iowa. USDA-SCS, Washington, DC.

Brubaker, S.C., Jones, A.J., Lewis, D.R., Frank, K., 1993. Soil properties associated with landscape position. Soil Sci. Soc. Am. J. 57, 235–239.

Colvin, T.S., 1990. Automated weighing and moisture sampling for a field-plot combine. Appl. Eng. Agric. 6, 713–714.

Colvin, T.S., Jaynes, D.B., Karlen, D.L., Laird, D.A., Ambuel, J.R., 1997. Yield variability within a central Iowa field. Trans. ASAE 40, 883–889.

Costanza, M.C., Afifi, A.A., 1979. Comparison of stopping rules in forward stepwise discriminant analysis. J. Am. Stat. Assoc. 74, 777–785.

Evans, R.O., Fausey, N.R., 1999. Effects of inadequate drainage on crop growth and yield. In: Skaggs, R.W., van Schilfgaarde, J. (Eds.), Agricultural Drainage. ASA, CSSA, SSSA, Madison, WI, pp. 13–54.

Ferguson, R.B., Hergert, G.W., Schepers, J.S., Gotway, C.A., Cahoon, J.E., Peterson, T.A., 2002. Site-specific nitrogen management of irrigated maize: yield and soil residual nitrate effects. Soil Sci. Soc. Am. J. 66, 553–644.

Fraisse, C.W., Sudduth, K.A., Kitchen, N.R., 2001. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. Trans. ASAE 44, 155–166.

Hair Jr., J.F., Anderson, R.E., Tatham, R.L., 1987. Multivariate Data Analysis with Readings, second ed. Macmillan Publishing Company, New York.

Halvorson, G.A., Doll, E.C., 1991. Topographic effects on spring wheat yields and water use. Soil Sci. Soc. Am. J. 55, 1680–1685.

Jaynes, D.B., Colvin, T.S., Ambuel, J., 1993. Soil type and crop yield determinations from ground conductivity surveys. In: Winter Meeting of ASAE, Chicago, IL, December 14–17, Paper No. 93-3552. ASAE, St. Joseph, MI.

Jaynes, D.B., Colvin, T.S., Ambuel, J., 1995a. Yield mapping by electromagnetic induction. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Site-Specific Management for Agricultural Systems. Proceedings of Second International Conference on Precision Agriculture. Minneapolis, MN, March 27–30, 1994. ASA, CSSA, SSSA, Madison, WI, pp. 383–394.

Jaynes, D.B., Novak, J.M., Moorman, T.B., Cambardella, C.A., 1995b. Estimating herbicide partition coefficients from electromagnetic induction measurements. J. Environ. Qual. 24, 36–41.

Jaynes, D.B., Kaspar, T.C., Colvin, T.S., 2002. Comparison of techniques for defining yield potential zones in an Iowa field. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Proceedings of Sixth International Conference on Precision Agriculture. Minneapolis, MN, July 14–17. ASA, CSSA, SSSA, Madison, WI.

Jaynes, D.B., Kaspar, T.C., Colvin, T.S., James, D.E., 2003. Cluster analysis of spatiotemporal corn yield patterns in an Iowa field. Agron. J. 95, 574–586.

Kachanoski, R.G., Gregorich, E.G., Van Wesenbeeck, I.J., 1988. Estimating spatial variations of soil water content using noncontacting electromagnetic inductive methods. Can. J. Soil Sci. 68, 715–722.

Karlen, D.L., Colvin, T.S., 1992. Alternative farming system effects on profile nitrogen concentrations on two Iowa farms. Soil Sci. Soc. Am. J. 56, 1249–1256.

Kaspar, T.C., Colvin, T.S., Jaynes, D.B., Karlen, D.L., James, D.E., Meek, D.W., Pulido, V., Butler, H., 2003. Relationship between six years of corn yields and terrain attributes. Prec. Agric. 4, 87–101.

Kravchenko, A.N., Bullock, D.G., Boast, C.W., 2000. Joint multifractal analysis of crop yield and terrain slope. Agron. J. 92, 1279–1290.

Lark, R.M., 2001. Some tools for parsimonious modeling and interpretation of within-field variation of soil and crop systems. Soil Tillage Res. 58, 99–111.

Lark, R.M., Stafford, J.V., 1997. Classification as a first step in the interpretation of temporal and spatial variability of crop yield. Ann. Appl. Biol. 130, 111–121.

Logsdon, S., Prueger, J., Meek, D., Colvin, T., James, D., 1999. Crop yield variability as influenced by water in rain-fed agriculture. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Proceedings of the Fourth International Conference on Precision Agriculture. St. Paul, MN, July 19–22, 1998. ASA, CSSA, SSSA, Madison, WI, pp. 453–465.

Lund, E.C., Colin, P.E., Christy, D., Drummond, P.E., 1999. Applying soil electrical conductivity technology to precision agriculture. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Proceedings of the Fourth International Conference on Precision Agriculture. St. Paul, MN, July 19–22, 1998. ASA, CSSA, SSSA, Madison, WI, pp. 1089–1100.

McBride, R.A., Gordon, A.M., Shrive, S.C., 1990. Estimating forest soil quality from terrain measurements of apparent electrical conductivity. Soil Sci. Soc. Am. J. 54, 290–293.

Milligan, G.W., Cooper, M.C., 1985. An examination of procedures for determining the number of clusters in a data set. Psychometrika 50, 159–179.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17.

Olson, K.R., Nizeyimana, E., 1988. Effects of soil erosion on corn yields of seven Illinois soils. J. Prod. Agric. 1, 13–19.

Pennock, D.J., de Jong, E., 1987. The influence of slope curvature on soil erosion and deposition in hummock terrain. Soil Sci. 144, 209–217.

Rhoades, J.D., Corwin, D.L., 1981. Determining soil electrical conductivity—depth relations using an inductive electromagnetic soil conductivity meter. Soil Sci. Soc. Am. J. 45, 255–260.

SAS Institute Inc., 2000. SAS Online Document, Version 8. SAS Institute Inc., Cary, NC.

Sawada, M., 1999. ROOKCASE: an Excel 97/2000 Visual Basic (VB) add-in for exploring global and local spatial autocorrelation. Bull. Ecol. Soc. Am. 80, 231–234.

Stafford, J.V., Lark, R.M., Bolam, H.C., 1999. Using yield maps to regionalize fields into potential management units. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), Proceedings of the Fourth International Conference on Precision Agriculture. St. Paul, MN, July 19–22, 1998. ASA, CSSA, SSSA, Madison, WI, pp. 225–237.

Steinwand, A.L., Fenton, T.E., 1995. Landscape evolution and shallow groundwater hydrology of a till landscape in central Iowa. Soil Sci. Soc. Am. J. 59, 1370–1377.

Thompson, A.L., Gantzer, C.J., Anderson, S.H., 1991. Topsoil depth, fertility, water management, and weather influences on yield. Soil Sci. Soc. Am. J. 55, 185–191.

Timlin, D.J., Pachepsky, Ya., Snyder, V.A., Bryant, R.B., 1998. Spatial and temporal variability of corn grain yield on a hillslope. Soil Sci. Soc. Am. J. 62, 764–773.

Upton, G.J.G., Fingleton, B., 1985. Spatial Data Analysis by Example, vol. 1, Point Pattern and Quantitative Data. Wiley, New York, NY.

Van Alphen, B.J., Stoorvogel, J.J., 2000. A functional approach to soil characterization in support of precision agriculture. Soil Sci. Soc. Am. J. 64, 1706–1713.

Wibawa, W.D., Dludlu, D.L., Swenson, F.J., Hopkins, D.G., Dahnke, W.C., 1993. Variable fertilizer application based on yield goal, soil fertility, and soil map unit. J. Prod. Agric. 6, 255–261.

Williams, B.G., Hoey, D., 1987. The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. Aust. J. Soil Res. 25, 21–27.

Yang, C., Peterson, C.L., Shropshire, G.J., Otawa, T., 1998. Spatial variability of field topography and wheat yield in the palouse region of the pacific northwest. Trans. ASAE 41, 17–27.