Delineation of Site–Specific Management Zones by Unsupervised Classification of Topographic Attributes and Soil Electrical Conductivity

C. W. Fraisse, K. A. Sudduth, N. R. Kitchen

Abstract. *The objective of this research was to determine if unsupervised classification of topographic attributes and soil electrical conductivity could identify management zones for use in precision agriculture. Data collected in two fields located in central Missouri were used to test the proposed methodology. Principal component analysis was used to determine which layers of data were most important for representing within–field variability. Unsupervised clustering algorithms implemented in geographic information system (GIS) software were then used to divide the fields into potential management zones. Grain yield data obtained using a full–size combine equipped with a commercial yield sensing system and global positioning system (GPS) receiver were used to analyze the "goodness" of the potential management zones defined for each field. Principal component analysis of input variables for Field 1 indicated that elevation and bulk soil electrical conductivity (EC) were more important attributes than slope and Compound Topographic Index (CTI) for defining claypan soil management zones. The optimum number of zones to use when dividing a field may vary from year to year and was mainly a function of weather and the crop planted. The number of zones decreased if adequate moisture conditions were present throughout the cropping season (unpredictable) or if crops tolerant to water stress were planted (predictable). This classification procedure is fast, can be easily automated in commercially available GIS software, and has considerable advantages when compared to other methods for delineating within–field management zones.*

Keywords. *ARC/INFO, DEM, GIS, Precision farming, Site–specific management.*

Site–specific management, or precision agriculture, has the potential to change the way fields are managed through variable–rate application of inputs such as fertilizers, lime, seeds, and pesticides. A potential failing of site–specific management is that the decision rules for varying inputs commonly use recommendation algorithms originally developed for whole–field management. These algorithms were generally based on data obtained from multiple locations over large geographic areas. Development of agronomic strategies specific to areas of a field that are subject to unique combinations of potential yield–limiting factors would allow more accurate management of inputs. The task of determining sub–field areas is difficult due to the complex combination of factors that affect crop yields.

Three basic approaches have been used to delineate soil management zones for site–specific management (Bell et al., 1995). The first uses county (order II) soil surveys prepared by the National Cooperative Soil Survey program, which describe soil variability at scales typically ranging from 1:12,000 to 1:24,000. More detailed surveys (order I) have occasionally been used and provide soil information at approximately the 1:5,000 scale. Traditional soil surveys give a general understanding of the effects of soil mapping units on crop productivity. However, they were not intended for making within–field recommendations at the same scale used today for site–specific management (Mausbach et al., 1993). A second approach uses geostatistical interpolation techniques to estimate the spatial distribution of soil properties from a network of point measurements. A disadvantage of this method is the large number of soil samples that must be collected and analyzed for correct representation of the variability present (Wollenhaupt et al., 1997).

A third approach uses temporally–stable data, such as bulk soil electrical conductivity (EC) and landscape features, to estimate patterns of soil variability using soil–landscape models. Landscape position and topographic attributes have been widely used to map within–field areas of high and low productivity based on water availability (Jones et al., 1989; Mulla et al., 1992; Jaynes et al., 1995; Sudduth et al., 1997). In these studies, footslope positions generally out–yielded up–slope positions unless poor drainage caused ponding or lack of aeration. Soil EC has also been used to investigate yield variability caused by soil water differences (Jaynes et al., 1995; Sudduth et al., 1995).

Kitchen et al. (1998) compared the use of traditional soil surveys and a map overlay approach based on topsoil depth and elevation to delineate management zones. They concluded that the map overlay approach has the advantage of being based on georeferenced measurements that are repeatable, unlike traditional soil surveys. However, the delineation of zones based on the map overlay approach is

Article was submitted for review in May 1999; approved for publication by the Information & Electrical Technologies Division of ASAE in August 2000.

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 the USDA implies no approval of the product to the exclusion of others that may also be suitable.

The authors are **Clyde W. Fraisse,** ***ASAE Member Engineer,*** former Agricultural Engineer, **Kenneth A. Sudduth,** ***ASAE Member Engineer,*** Agricultural Engineer, and **Newell R. Kitchen,** Soil Scientist, USDA–ARS, Cropping Systems and Water Quality Research Unit, Columbia, Missouri. **Corresponding author:** Kenneth A. Sudduth, Agricultural Engineer, USDA–ARS–CSWQRU, 269 Agricultural Engineering Bldg., Univ. of Missouri, Columbia, MO 65211; phone: 573–882–1114; fax: 573–882–1115; e–mail: SudduthK@missouri.edu.

dependent on arbitrary classification criteria defined by the user, which include the number of classes and class breaks defined for each variable. Normally, the map layers representing each variable are first reclassed in a user–defined number of classes. Following that, the resulting maps are overlaid, and the unique combinations of classes define the management zones in the field. If the user defines a different number of classes for a given variable, then the resulting classification will be different, and consequently, different management zones will be produced.

An additional approach is the use of unsupervised clustering algorithms, which are available in most commercial Geographic Information Systems (GIS) software, for grouping similar areas in the field. This approach is fast, can be easily automated, and does not require the initial reclassification of variables. It also allows the inclusion of additional layers that might be important to characterize the variability observed in the field, such as remote sensing images or yield maps, further refining or validating the delineation process.

The objective of this study was to develop automated, unbiased procedures for defining sub–field management zones through the application of unsupervised clustering algorithms to soil EC measurements and topographic attributes calculated from elevation measurements.

Materials and Methods

Research Fields Description

Elevation and soil EC data were collected on two fields, 36 ha (Field 1) and 28 ha (Field 2) in size, located near Centralia, Missouri. The soils of the study fields are characterized as claypan soils (fine, smectitic, mesic, Aeric Vertic Epiaqualfs and Vertic Albaqualfs). These soils are poorly drained and have a restrictive, high–clay content layer (the claypan) occurring below the topsoil. Detailed elevation data were obtained using a total station surveying instrument and standard mapping procedures. Field 1 elevation ranges from 265.8 m at the southeast corner to 261.9 m at the drainage outlet along the north edge of the field (fig. 1). Surface and subsurface water flows from the west and east sides of the field to a central natural drainage channel that carries the water to the outlet. Field 2 elevation ranges from 264.5 m at the southeast boundary to 261.0 m at the drainage outlet in the northwest corner of the field.

Topographic attributes were calculated from elevation data using TAPES–G (Terrain Analysis Program for the Environmental Sciences–Grid version), a program designed primarily to calculate hydrological factors from a Digital Elevation Model (DEM)(Gallant and Wilson, 1996). Attributes calculated were slope, profile curvature, tangential curvature, and compound topographic index (CTI), since each of these was hypothesized to be useful in differentiating areas of the field where yield–limiting factors varied. Slopes were obtained using the D8 approach, which calculates gradient as the steepest slope from the central node to one of its eight nearest neighbors. Profile curvature measures the rate of change in the direction of maximum slope. Tangential curvature is the curvature in the normal plane in the direction perpendicular to the gradient (or tangent to the contour line). Tangent curvature can be expected to provide a good representation of the collection or dispersal of water flowing over a surface. The CTI or wetness index (eq. 1) is a function of specific catchment area (A_S) and the slope gradient (β), and represents the spatial distribution of water accumulation areas in the landscape.

$$\mathrm{CTI} = \ln \frac{A_S}{\tan\beta} \qquad (1)$$

Figure 2 shows slope and CTI variability for Field 1. The darker areas of the CTI map indicate zones of surface saturation or higher soil moisture content.

Electrical conductivity of the soil was measured using an EM38 (Geonics Limited, Mississauga, Ontario, Canada). The EM38 uses the principle of electromagnetic induction to quantify soil EC in milliSiemens per meter (mS/m). The instrument was operated in the vertical dipole mode, providing an effective measurement depth of approximately 1.5 m (McNeil, 1992). Measurements in the field were performed using a mobile system that included an all–terrain vehicle, a wooden trailer for carrying the EM38, a DGPS receiver, and a computer for data acquisition (Kitchen et al., 1996). Previous work (Sudduth et al., 1995) demonstrated that areas of high soil EC readings correspond to shallow topsoil (<20 cm) where the claypan horizon is closer to the soil surface. Low readings are observed in areas of the field with deep topsoil (>60 cm), generally located at lower elevations where there is deposition of eroded topsoil material. Figure 3 is a soil conductivity map of Field 1.

Yield Data

Grain yield data were obtained using a full–size combine equipped with a commercial yield sensing system and global positioning system (GPS) receiver. Data for Field 1 were obtained for corn (1993, 1997), soybean (1994, 1996), and grain sorghum (1995) crops. Data for Field 2 were obtained for corn (1996) and soybean (1995, 1997) crops. Yield data were analyzed using geostatistics, and appropriate semi–variogram models and parameters were used to krige the data to a grid with a 10–m cell size.

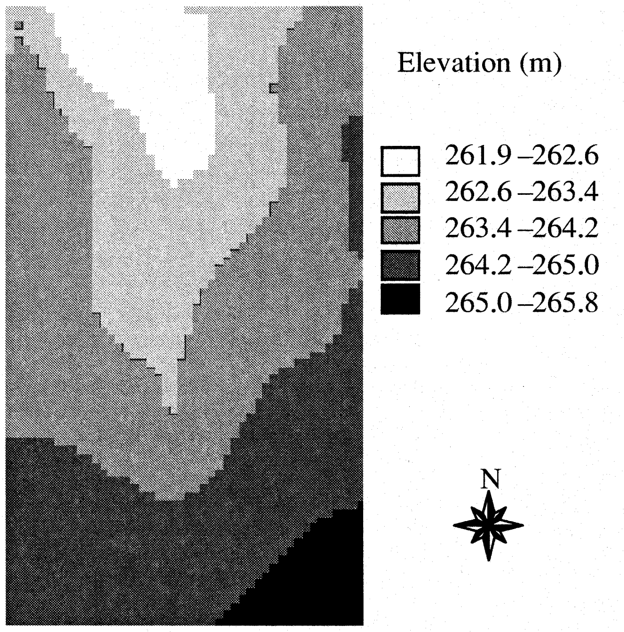

Figure 1. Elevation map of Field 1.

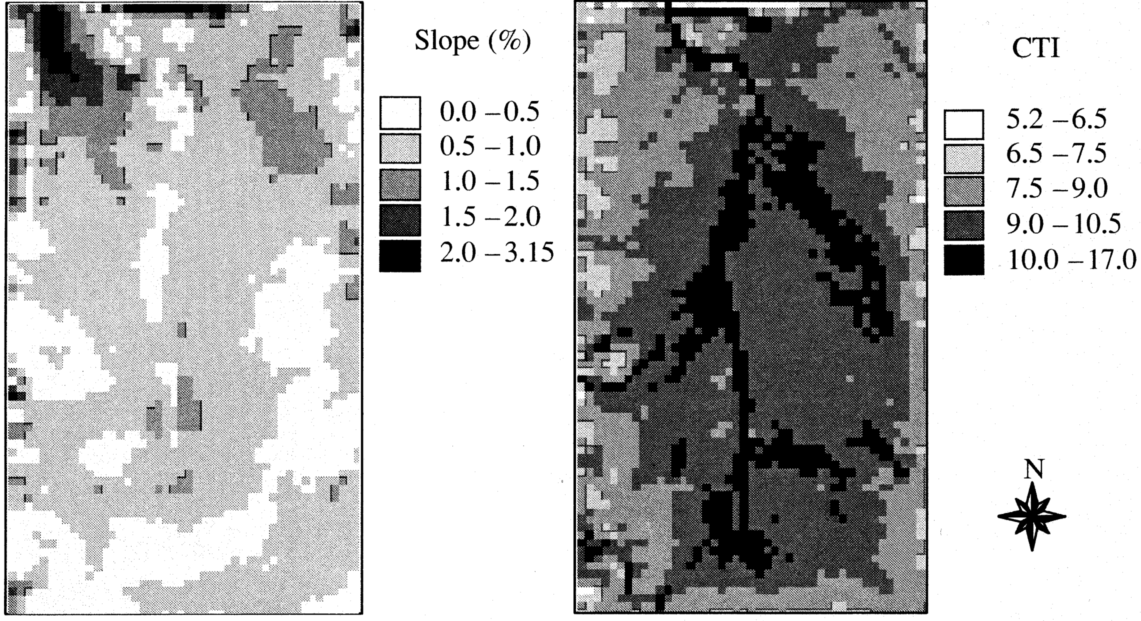

Figure 2. Slope and compound topographic index (CTI) maps of Field 1.

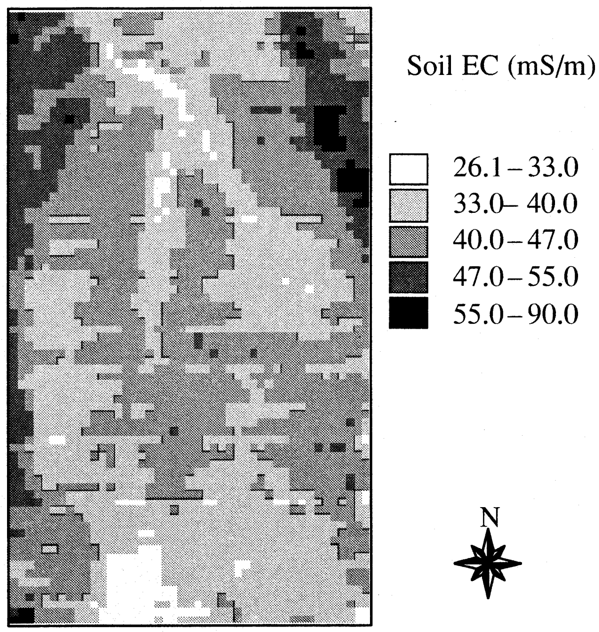

Figure 3. Soil electrical conductivity map of Field 1.

Principal Component Analysis

Principal Component Analysis (PCA) is a classical statistical technique that linearly transforms an original data set of variables. The general purpose of PCA is data description and interpretation. In practice, PCA reduces the dimensionality of problems and transforms interdependent variables into significant and independent ones. This linear transformation can compress the original data set into a substantially smaller set of uncorrelated variables, the principal components, that represents most of the information in the original set of variables (Dunteman, 1989). The most important variables describing the total variation in a data set can also be selected by PCA. That is, rather than substituting the principal components for the original variables, we can select a set of variables that have high correlation with the major principal components.

In this study, PCA was used to help determine which variables were most important for the characterization of variability and should therefore be used in the unsupervised classification. The main goal in this process was to reduce the number of variables without losing important information. We also wanted to know if the measurement of soil EC was necessary or if successful classification was possible using only topographic attributes. The Unscrambler, version 5.5, software (Computer Aided Modeling, 1994) was used for describing and analyzing the data set by PCA.

Unsupervised Classification

The main objective in an unsupervised classification or clustering is to identify naturally occurring clusters in the data. The unsupervised clustering algorithm used was the ISODATA (Iterative Self–Organizing Data Analysis Technique) procedure (Tou and Gonzalez, 1974). This algorithm is commonly used for satellite image classification based on spectral reflectances from multiple wavebands (Irvin, 1996). It was accomplished using the ISOCLUSTER function in the Arc/Info GIS software, Grid module (ESRI, 1994). The ISOCLUSTER function uses a modified iterative optimization procedure, also known as the migrating means technique. The process starts with arbitrary means being assigned by the software, one for each cluster (the number of clusters is dictated as a user input). Every cell is then assigned to the closest of these means, all in the multidimensional attribute space. New means are then recalculated for each cluster based on the attribute distances of the cells that belong to the cluster after the first iteration. The process is repeated enough times to ensure that the migration of cells from one cluster to another is minimal and that all the clusters become stable. The user specifies the number of classes, number of iterations, minimum number of cells in a class, and sampling interval. The ISOCLUSTER function returns a signature file containing class means and covariance matrices, which are then used as input for the maximum likelihood classifier (the MLCLASSIFY function in Arc/Info). The classifier uses the mean vector and covariance matrix of each class to compute the statistical probability that a grid cell belongs to a class. Each cell is assigned to the class for which it has the highest probability of being a member.

In order to characterize resulting classes by mean vectors and covariance matrices, the data for each class should have a roughly Gaussian distribution. Elevation, profile curvature, tangential curvature, and soil EC all exhibited adequate distributions (fig. 4). However, the slope data histogram was slightly skewed and was therefore log transformed to improve its distribution. The distribution of CTI was also skewed, with the higher values separated from the main distribution representing the drainage channel where the specific catchment area (A_s) is larger and the slope gradient (β) is generally lower. This data was also log transformed, but this provided only a modest improvement in approximating a normal distribution.

Better results are obtained in the clustering process if all layers in the stack have the same data ranges. The data ranges in the original data were very different. Therefore, all grids were transformed using equation 2 to a range between 0 and 100 prior to PCA analysis and ISOCLUSTER classification.

$$Z = \frac{(X - old_{min}) \times (new_{max} - new_{min})}{old_{max} - old_{min}} + new_{min} \qquad (2)$$

where

Z = output grid with new data ranges

X = input grid

old_{min} = minimum value of the input grid

old_{max} = maximum value of the input grid

new_{min} = desired minimum value for the output grid (0)

new_{max} = desired maximum value for the output grid (100)

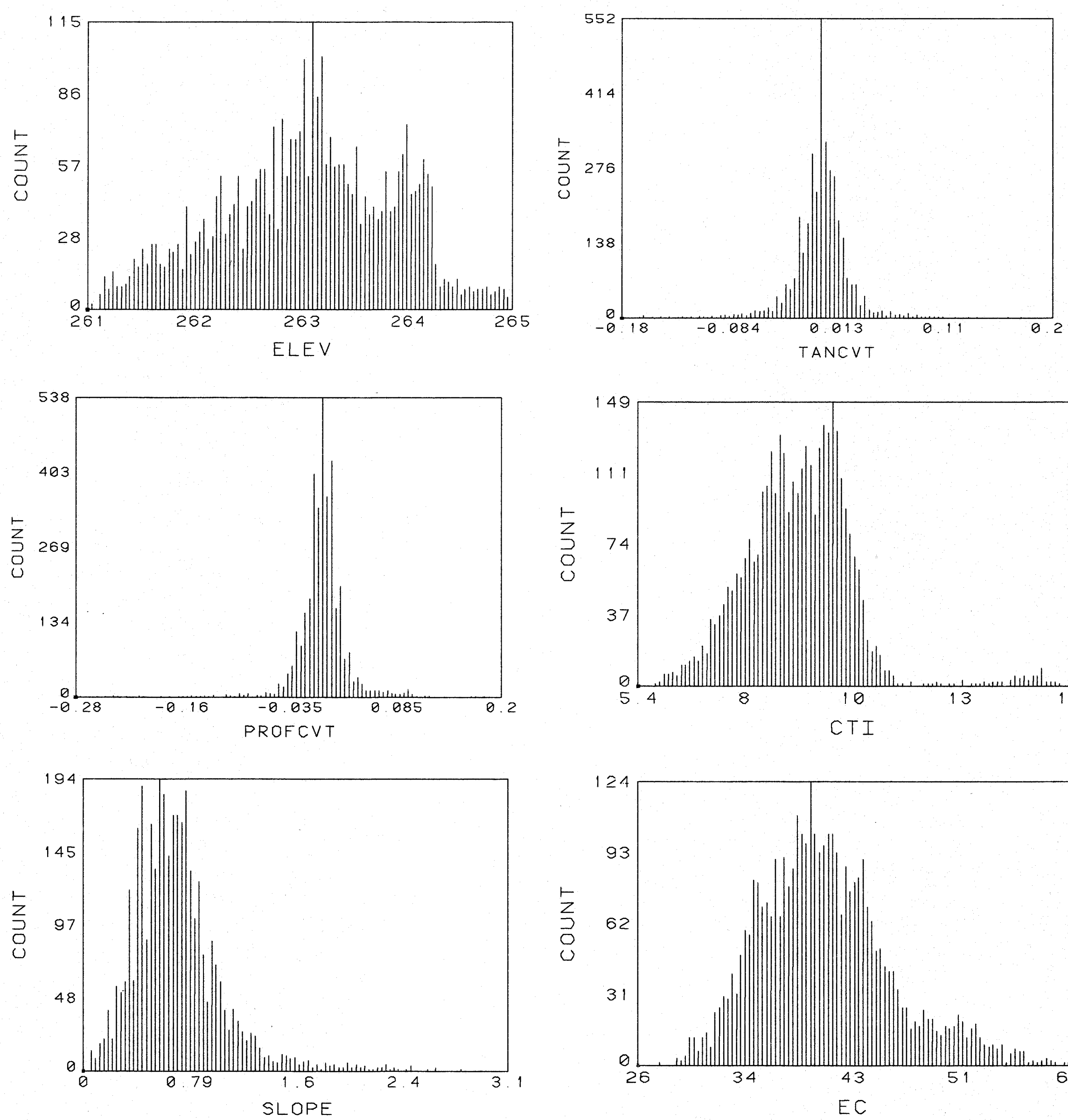

Figure 4. Histograms of the variables (elevation, tangential curvature, profile curvature, CTI, slope, and soil EC) used for principal component analysis of Field 1 data.

Optimum Number of Zones Based on Yield Analysis

Selection of variables to be used in the unsupervised classification is the first step in the process of delineating within–field management zones. The classification procedure itself determines the boundaries of the zones based on the spatial structure of the input variables, so no user intervention is required in that step. However, the user must somehow determine the correct or optimum number of zones to create. It is logical to assume that evaluation of the "goodness" of a zone should be based on an independent mapped variable that integrates the effects of limiting factors. Remotely sensed vegetative index maps or productivity maps might be used to provide this independent evaluation. However, in this study measured grain yield was chosen as the evaluation variable.

The unsupervised classification procedure was used to divide both fields into 2, 3, 4, 5, and 6 management zones. The zones were created using elevation, slope, and soil EC as input layers. Yield data for Field 1 (1993–1997) and Field 2 (1995–1997) were normalized by dividing the value measured for each grid cell by the mean value for the entire field. Yield statistics were calculated for each zone and year in all maps. The main goal of this procedure was to determine how much was gained in terms of yield "uniformity" within a zone by dividing the field into additional zones. If we kept dividing the field into a larger number of zones, ultimately there would be as many zones as grid cells and no yield variability within a zone. Yield variances for each zone were weighted based on the area of the zone as follows:

$$S_Z^2 = \frac{1}{n_Z} \sum_{i=1}^{n_Z} (Y_i - m)^2 \times \frac{n_Z}{n_T} \quad (3)$$

where

S^2_z = weighted variance for zone Z

Y_i = yield measured for cell i

m = mean of measured yield in zone Z

n_Z = number of cells in zone Z

n_T = total number of cells in the map

The decrease in within–zone yield variance was used to select the most appropriate number of zones. Total within–zone yield variance of a map (S_T^2) was defined as the sum of weighted within–zone yield variances for each zone, as follows:

$$S_T^2 = S_1^2 + S_2^2 + \ldots + S_Z^2 \quad (4)$$

Results and Discussion

Yield patterns varied considerably from year to year due to climate and crop differences. The 1993 growing season was characterized by frequent and heavy rains, with total precipitation of 157 cm. In lower parts of the landscape, excess water reduced yields. The 1994 precipitation of 82 cm was below the 58–year average of 92 cm, and the 5 cm of rainfall received in July and August caused drought stress, which reduced yields. In 1995, precipitation was 115 cm, with an excessively wet planting season which again caused stand problems and some yield reductions in the lower portion of the landscape. In 1996, there was adequate within–season precipitation (60 cm), and measured yields across Field 1 were more uniform than yields for other years. The 1997 growing season was characterized by a wet spring followed by mid–season drought stress during the pollination period, which reduced corn yields in areas with eroded, shallow topsoil.

Principal Component Analysis

Principal component analysis was performed for two sets of Field 1 data. The variables included in the first data set were elevation, slope, profile curvature, tangent curvature, CTI, and soil EC. The second data set included the same variables except for CTI. Removing CTI was done to test how important this variable was in explaining the variability observed in the field. Most of the information contained in the CTI map was very similar to that provided by the soil EC data. Areas with high CTI values were also areas with deep topsoil and, consequently, lower soil EC. Areas with deeper topsoil were areas where eroded material accumulated and therefore were normally located in the lower and flatter areas of the field. These areas were also subject to water accumulation and higher soil moisture conditions characterized by higher CTI values.

Table 1 shows the percentage of the total data set variance that could be explained by each principal component in the analysis. The first principal component always explains the largest fraction of the total data set variance. The second component explains the next–largest fraction, and so on. When the number of components equals the number of original variables, 100% of the variance in the original dataset is explained. In the case of the first data set with six variables, the first and second principal components together explained 63.1% of the variance, and the first four principal components explained 87.2% of the variance. In the case of the second data set, with five variables, the first principal component explained 41.9% of the variance, and the first and second principal components together explained 70.6% of the variance.

Table 2 shows the loadings of the principal components of each data set. Loadings express the relationship between input variables and the principal components and indicate which variables are influencing the model. The most important variables are those that exhibit large loadings in those principal components that explain most of the variance in the data. It is clear that elevation was an important variable in both data sets, having loadings of 0.94 and 0.93 in the first principal component of data sets 1 and 2, respectively. Soil EC was also important, with loadings of 0.77 and 0.93 in the second principal component of data sets 1 and 2, respectively. Slope had loadings of 0.55 and 0.93 in the third principal component of data sets 1 and 2, respectively. CTI had loadings of –0.15, –0.53, and –0.54 in the first, second, and third principal components of data set 1. Relative to elevation, soil EC, and slope, CTI can be considered as a variable of moderate importance for explaining the variability found in Field 1. Neither profile nor tangent curvature contributed much to the main principal components.

Figures 5 and 6 show maps of principal components 1 and 2 of both data sets. The similarity of PC 1 to the elevation map of the area (fig. 1) demonstrates that elevation is a dominant variable in both data sets. Soil EC is dominant in PC 2, but the effect of CTI can also be noticed in data set 1. The drainage channel that corresponds to areas of high CTI can be noticed in the PC 2 map (fig. 5).

Table 1. Individual and cumulative percentage of total data set variance explained by each principal component (PC).

Data Set	Principal Component	Variance Explained (%) Individual	Cumulative
Data set 1 (elevation, slope, profile curvature, tangential curvature, CTI, and soil EC)	PC1	34.4	34.4
	PC2	28.7	63.1
	PC3	12.3	75.4
	PC4	11.8	87.2
	PC5	4.2	91.4
	PC6	8.6	100.0
Data set 2 (elevation, slope, profile curvature, tangential curvature, and soil EC)	PC1	41.9	41.9
	PC2	28.7	70.6
	PC3	14.0	84.6
	PC4	6.8	91.4
	PC5	8.6	100.0

Table 2. Variable loadings in the principal components calculated for data sets 1 and 2.

Variables		Elevation	Slope	CTI	Profile Curvature	Tangent Curvature	Soil EC
Data Set 1	PC1	0.94	–0.27	–0.15	0.03	0.04	–0.15
	PC2	0.12	0.31	–0.53	0.08	0.10	0.77
	PC3	–0.03	0.55	–0.54	0.07	0.15	–0.61
	PC4	0.32	0.71	0.51	–0.12	–0.33	0.08
	PC5	0.05	0.13	0.36	0.62	0.69	0.04
	PC6	–0.04	–0.07	–0.01	0.77	–0.62	–0.03
Data Set 2	PC1	0.93	–0.30	n.a.	0.02	0.03	–0.21
	PC2	0.28	0.23	n.a.	0.06	0.06	0.93
	PC3	0.23	0.93	n.a.	0.03	–0.01	–0.30
	PC4	–0.06	–0.01	n.a.	0.53	0.84	–0.07
	PC5	–0.02	–0.03	n.a.	0.85	–0.53	–0.01

The high loadings of elevation and soil EC in the main principal components demonstrated that, at a minimum, they should be considered in the unsupervised classification procedure. Slope and CTI should also be taken into consideration but may have less effect on the delineation of the zones. In fact, CTI is more relevant to the analysis if the field being divided into zones has a pronounced drainage pattern, as is the case for Field 1. Because loadings from profile and tangential curvatures were small in the main principal components, they could be excluded from the unsupervised classification process.

Unsupervised Classification

The unsupervised classification procedure was used to delineate management zones in Fields 1 and 2, starting with 2 zones and further dividing the fields into a maximum of 7 zones. Less pronounced and possibly less interpretable features were included as the number of zones increased. Figure 7 shows the results obtained when Field 1 was classified into 3 and 6 zones using 6 variables (elevation, slope, profile curvature, tangential curvature, CTI, and soil EC) for the clustering algorithm. The 3–zone classification roughly follows the elevation and soil EC patterns of the field (figs. 1 and 3). Zone 1 included the lower elevation area in the north central portion of the field. Zone 2 included the upland areas located in the southern half of the field. Zone 3 included the sloping and shallow soils (high soil EC) located along the east and west sides of the field.

In the 6–zone classification, the results can be interpreted as a combination of elevation and soil EC layers. Zone 1 included the low elevation areas of the field near the drainage channel and with deeper topsoil (low soil EC). Zone 2 corresponded to the side slopes that drain to the central channel. Zone 3 included areas with medium elevations and soil EC values. Zone 4 included areas of the field with gentle slopes but relatively shallow topsoil (above average soil EC). Zone 5 included the southern area of the field with gentle slopes, higher elevation, and deep topsoil (low soil EC). Zone 6 included the shallow eroded soils adjacent to the east and west boundaries of the field.

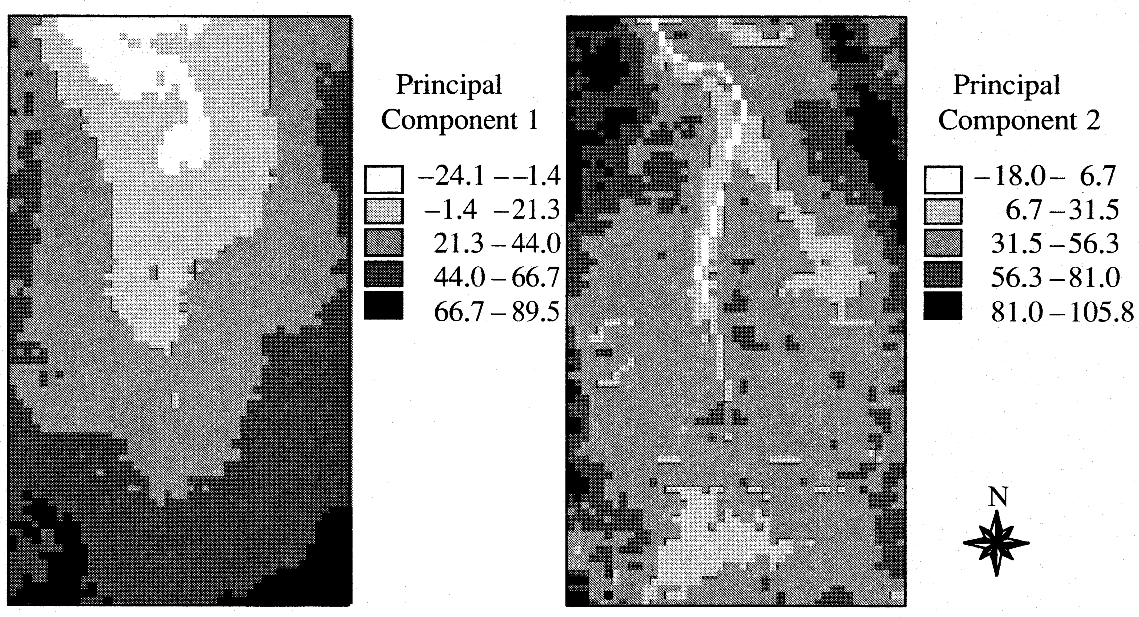

Figure 5. Principal components 1 and 2 for Field 1, derived from data set 1 (elevation, slope, profile curvature, tangential curvature, CTI, and soil EC).

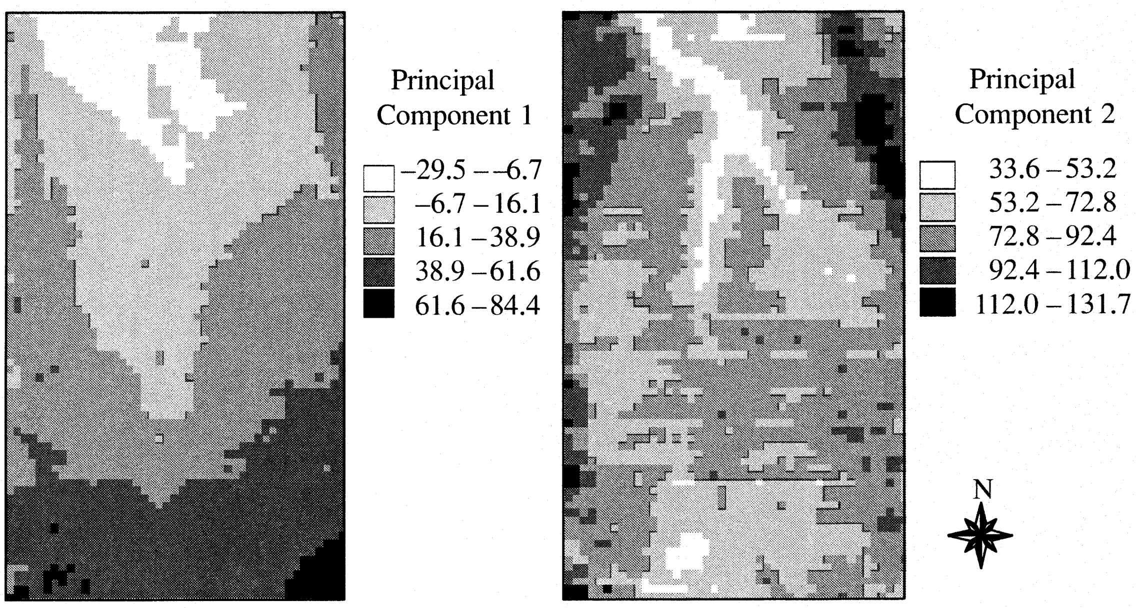

Figure 6. Principal components 1 and 2 for Field 1, derived from data set 2 (elevation, slope, profile curvature, tangential curvature, and soil EC).

Figure 8 shows the results obtained for classification of Field 1 into 5 zones using 6 variables (elevation, slope, CTI, profile curvature, tangential curvature, and soil EC), 4 variables (elevation, slope, CTI, and soil EC), and 3 variables (elevation, slope, and soil EC). The resulting maps confirmed the results obtained with the PCA, indicating that removing the profile and tangent curvatures from the classification process makes little difference in the delineation of the zones. The zones obtained by clustering 6 and 4 variables were basically the same. The results also indicated that removing CTI and using only elevation, slope, and soil EC in the classification process also resulted in very similar zones.

The results obtained from the automated process of unsupervised classification using topographic attributes and soil EC were somewhat similar to the Order 1 soil surveys conducted in Field 1 on three different occasions by the USDA Natural Resource Conservation Service (NRCS). In 1991, the Missouri State office of the USDA–NRCS conducted an Order 1 survey (ISS91) at a 1:5,000 scale. In 1993, a revised order 1 soil survey (ISS93) was done with more laboratory data analysis. Then in 1997, a team of soil scientists from the Missouri State NRCS office and the National Soil Survey Center in Lincoln, Nebraska, conducted a third Order 1 soil survey (ISS97). The three surveys included some similar patterns but were very different in their level of detail, the soil types included, and the location of the boundaries between soil types. Figure 9 illustrates the lack of repeatability that can be encountered when using subjective classification methods to generate soil zone boundaries. By contrast, results obtained with an objective classification method, such as the one proposed here, would be much more repeatable.

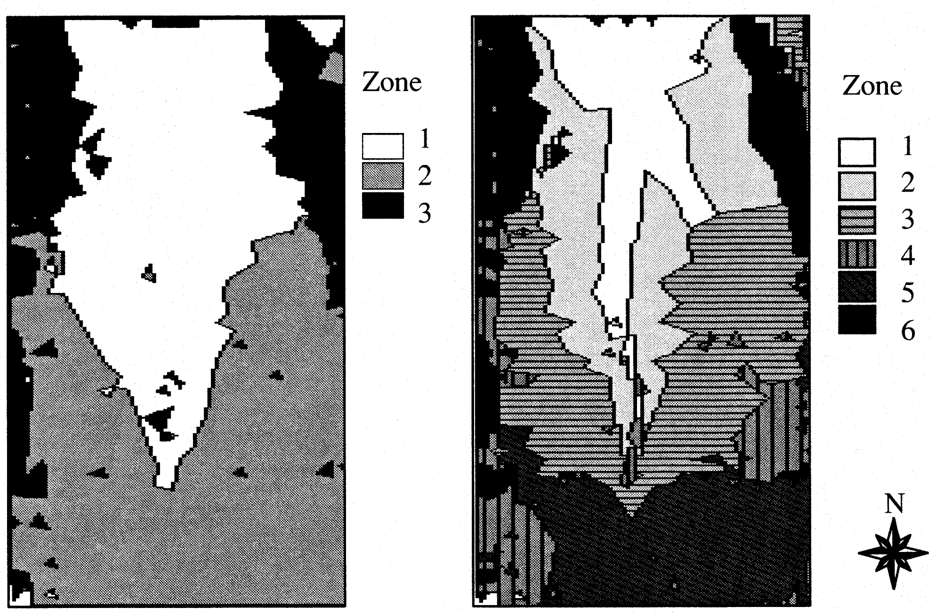

Figure 7. Classification of Field 1 into 3 (left) and 6 (right) zones using 6 variables: elevation, slope, profile curvature, tangential curvature, CTI, and soil EC.

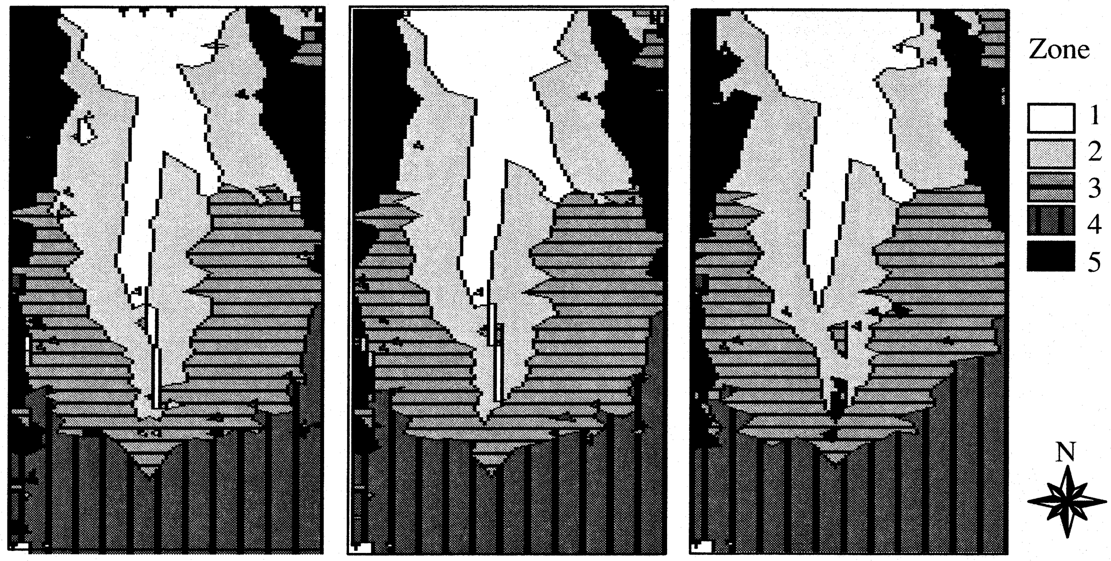

Figure 8. Classification of Field 1 into five zones using six variables: elevation, slope, profile curvature, tangential curvature, CTI, and soil EC (left); four variables: elevation, slope, CTI, and soil EC (center); and three variables: elevation, slope, and soil EC (right).

Figure 10 shows the results obtained for the classification of Field 1 into 5 zones using the map overlay approach described by Sudduth et al. (1996). The map overlay method divided the field into 5 sub–field areas on the basis of elevation and topsoil depth. The field was first divided into areas of low (<20 cm), medium, and high (>50 cm) topsoil depth. The medium and high topsoil depths were then sub–divided into the lower 1/3 of the landscape and the higher 2/3 of the landscape. The low topsoil and high topsoil zones—at low elevation—created by the map overlay approach (fig. 10) were similar to zones obtained with the unsupervised classification approach (fig. 8). However, the other three zones were defined differently by the two methods. The map overlay method placed over half of the field into one zone, while the unsupervised classification method divided the field into more equal sections.

Optimum Number of Zones Based on Yield Analysis

Yield data obtained for Field 1 (1993–1997) and Field 2 (1995–1997) were used to investigate the optimal number of zones to use when dividing a field. Figure 11 shows that within–zone yield variance decreased as Field 1 was divided into management zones. The yield variance for the entire field not divided into zones (i.e., one management zone) was used as the reference, or 100% level. In this case, the total variance in yield data is within the zone. As additional zones are implemented, an increasing part of the total yield variance is explained by the zone partitioning (between–zone variance), causing the total within–zone variance to decrease. In Field 1, a maximum decrease in yield variance was obtained by dividing the field into five management zones. Depending on the year, little reduction or even an increase in variance was obtained by further dividing the field into six management zones.

For years in which the crop was subjected to water stress (1994 and 1997), up to 32% of the yield variance was explained by five zones. Similar results were found for a year of adequate moisture (1996). In 1996, 26% of the yield variance was explained by dividing the field into two zones. Soybean yield in 1996 was more uniform due to adequate soil moisture throughout most of the cropping season, and the division of the field into two zones was enough to obtain a substantial decrease in yield variability within zones. In the case of 1994 and 1997, within–zone variance decreases of

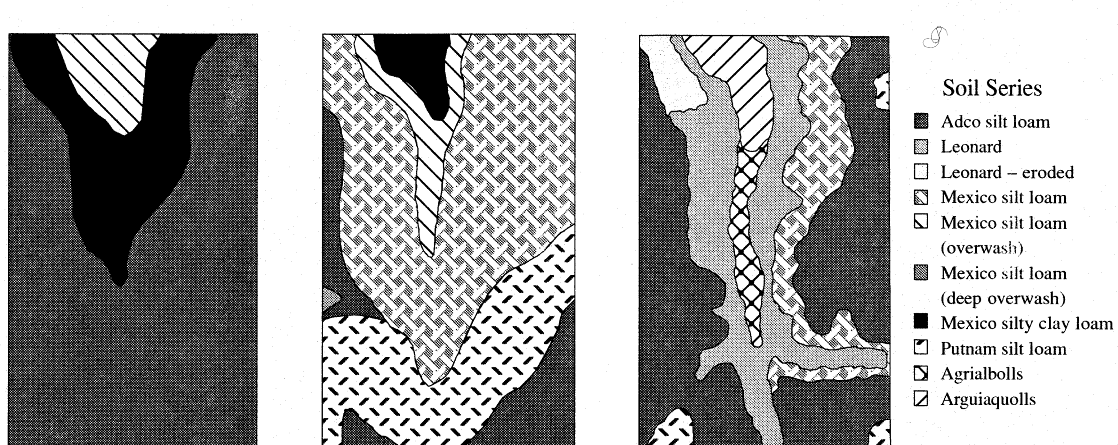

Figure 9. NRCS Order 1 soil surveys of Field 1: ISS91 (left), ISS93 (center), and ISS97 (right).

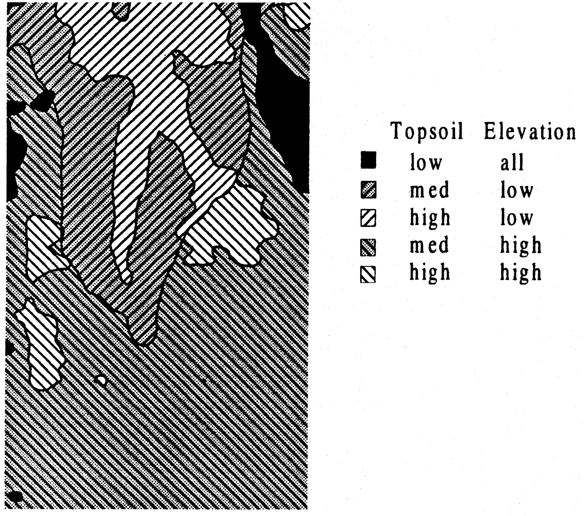

Figure 10. Zones defined for Field 1 by the map overlay method.

only about 6% were obtained by dividing the field into two management zones. Larger decreases for these years were obtained by further dividing the field into as many as five management zones.

Insufficient precipitation during critical pollination and seedfill periods caused more yield variability due to differences in water holding capacity associated with landscape position. In 1993 and 1995, the maximum yield variance explained by dividing the field into five management zones was 12% and 10%, respectively. Similarly to 1996, most of this decrease in within–zone variance was obtained by dividing the field into two zones. Above–average precipitation during the 1993 cropping season caused yield to be more uniform in the upper areas of the field, independent of topsoil depth, and to be reduced in the lower areas due to excess water.

The results obtained indicate that the optimum number of zones to use for dividing a field is a function of weather conditions and the crop planted. Division of Field 1 into five management zones (fig. 12) seems to be the most appropriate, especially during years in which water stress conditions are present and more drought–susceptible crops, such as corn, are planted. If adequate or excess moisture conditions occur, then management zones located in the upper areas of the field will probably behave similarly. Table 3 shows the mean elevation, soil EC, and slope values for each of the management zones defined for both fields.

Figure 13 shows the average normalized yield for the five zones of Field 1. Average normalized yield for zone 1, the lower elevation area of the field, was quite variable, ranging from 0.96 in 1993 to 1.24 in 1994. This result was not a surprise since the low areas of the field generally have lower yields during years of excessive precipitation, such as 1993, and high productivity during dry years, such as 1994. The average yields were very similar for the dry years of 1994 and 1997 throughout the various zones, except for zone 4, which had a lower average in 1994. Zone 4 is an area of flat topography located toward the south side of the field. It presented quite "stable" yields for all the other years used in the analysis, suggesting that, in 1994, a factor not considered in the classification procedure may have had a negative impact on yield in this zone. Zones 2 and 5 presented quite similar behavior, with lower average yields during dry years and "normal" (close to 1.0) average yields during years of excessive or adequate rainfall. Zones 2 and 5 are located on sloping shallow eroded soils, and lower yields during dry years are not unexpected. Zone 3 was more consistent in terms of yield, and for most of the years, its average normalized yields were close to 1.0.

Figure 14 shows the percent of yield variance explained in Field 2 by dividing the field into management zones. A maximum decrease in within–zone yield variance of 37% was obtained by dividing the field into four management zones. Most of the yield variance was explained by dividing the field into two zones. However, further division into four zones decreased the total variance by an additional 6% in 1996 and 1997.

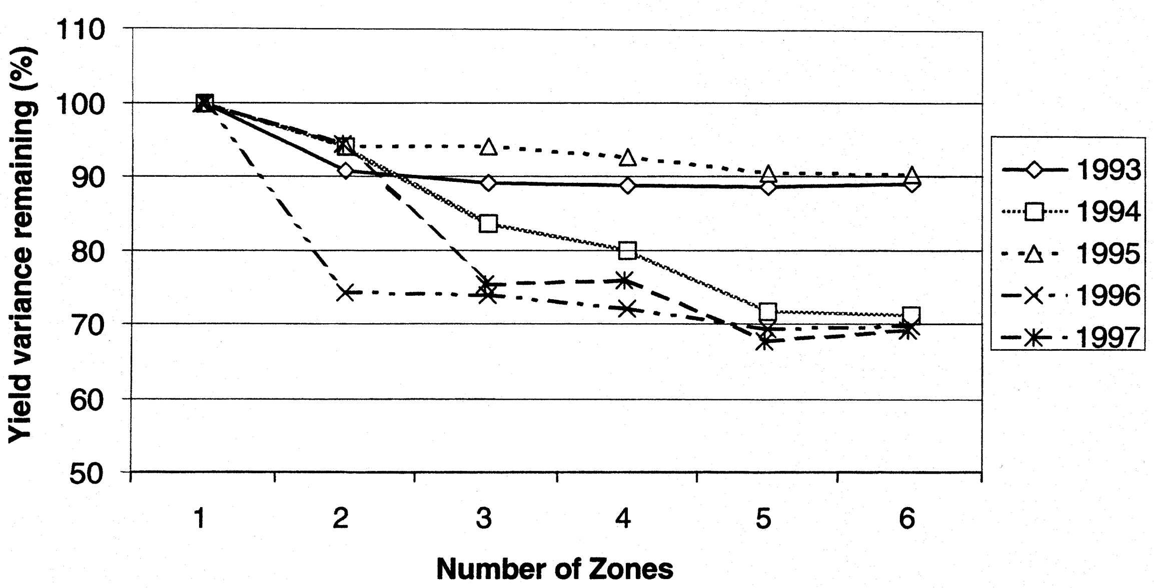

Figure 11. Portion of within–zone yield variance remaining after dividing Field 1 into management zones.

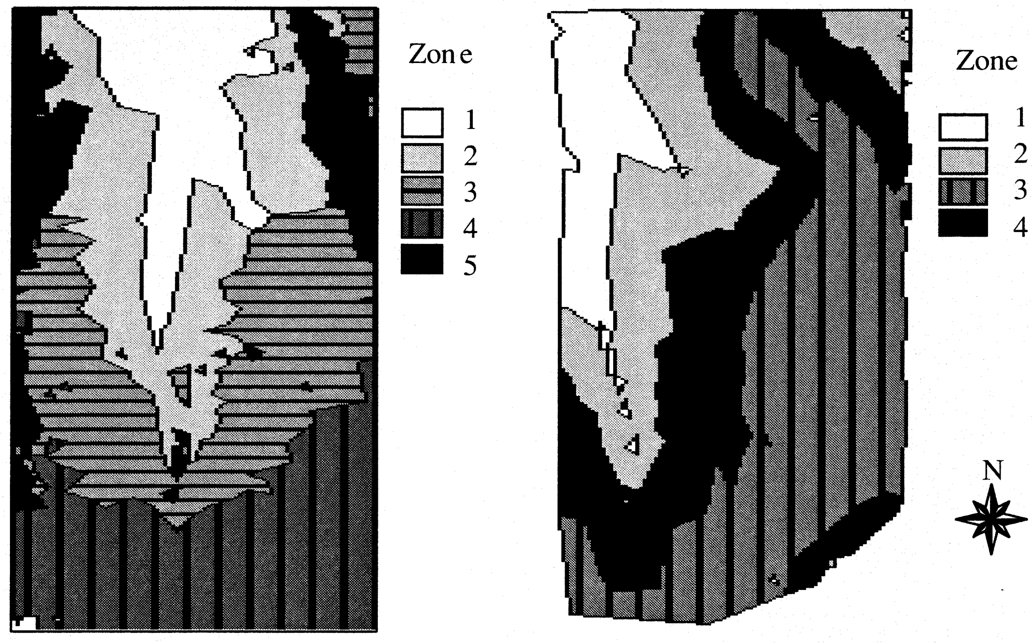

Figure 12. Optimum number of management zones determined for Field 1 (left) and Field 2 (right).

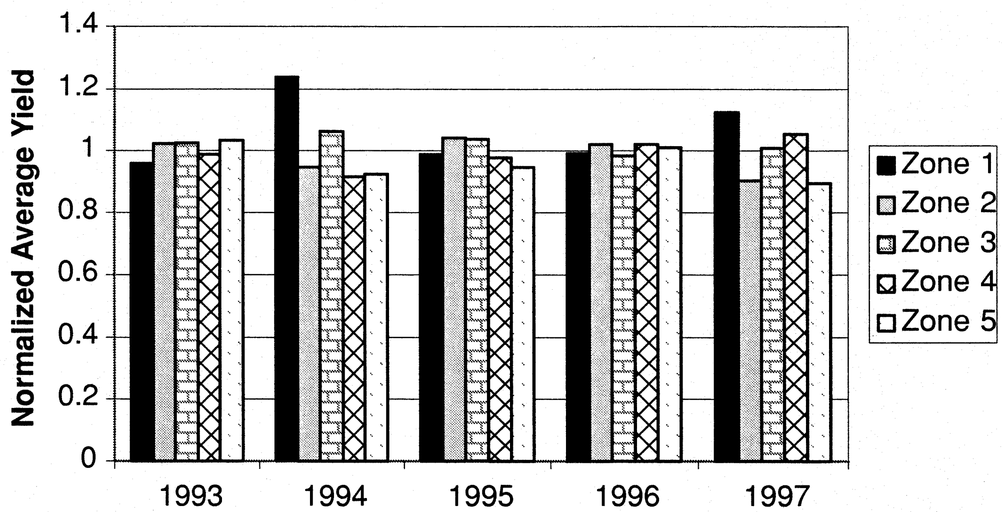

Figure 13. Calculated average normalized yield of five potential management zones defined for Field 1 (1993–1997).

Table 3. Elevation, soil EC, and slope of optimum management zones defined for Fields 1 and 2.

Field	Zone	Mean Elevation (m)	Mean Soil EC (mS/m)	Mean Slope (%)
1	1	262.6	37.1	0.74
	2	263.3	43.4	0.94
	3	264.0	39.7	0.57
	4	264.9	38.1	0.51
	5	263.9	50.9	0.88
2	1	261.5	30.0	0.22
	2	262.2	37.3	0.74
	3	263.8	32.5	0.31
	4	263.1	35.4	0.58

Figure 15 shows the average normalized yield for the four zones of Field 2. Comparison of figure 15 with figure 13 shows that the yield variation between zones was similar for the two fields in 1996 and 1997. However, in 1995, the between–zone yield differences were much higher for Field 2, due to differences in planting dates and crops grown. In that year, Field 2 was planted to soybean, and significant yield reductions were seen due to excess water in the lower elevation areas. Field 1 was planted to grain sorghum at a later date in 1995, so early–season excess moisture did not affect crop stands in that field.

Average yield for zone 1 (the lower elevation area) of Field 2 was variable, ranging from 0.74 in 1995 to 0.94 in 1996. In addition, average yields for this zone were always below the field average for the 3 years analyzed, which differs from the results obtained for the lower elevation areas of Field 1. The lower yield for zone 1 in 1995 can be explained by the above–average precipitation that year, which caused excess moisture conditions in the lower elevation areas of the field. However, the average yield for zone 1 was also lower in 1996 and 1997, despite the fact that deeper topsoil (lower soil EC) is found in that area of the field. In fact, most of the area included in zone 1 was replanted in 1997 due to emergence problems caused by a wet spring. This suggests that drainage should be considered as a valid management option to increase yields in zone 1.

Results obtained for zone 2 were quite similar to zone 1, with measured yields normally under the field average. Zone 2 is located at lower elevations and includes areas with shallow topsoil depths. Average yields for 1996 and 1997 were quite stable for all zones. However, in 1995, the excess precipitation in early spring decreased yields in the lower parts of the field.

Zone 3 is the highest producing area of Field 2, with normalized average yields ranging from 1.07 in 1996 to 1.13 in 1995. Zone 3 is located in the upper areas of the field with medium and high topsoil depths (medium to low soil EC). The normalized average yields for zone 4 were 1.1 in 1995 and slightly under 1.0 for 1996 and 1997. Zone 4 does not present major drainage problems, but it includes areas with shallow topsoil depths (higher soil EC readings).

Zones created by the unsupervised classification procedure were compared to the zones from the NRCS Order 1 surveys and the map overlay classification on the basis of yield variance (table 4). In almost all cases, the zones defined by the optimum unsupervised classification offered a better explanation of yield variation than the zones created by the other approaches. The unsupervised classification procedure was generally better at defining soil zone boundaries on Field 1 than the map overlay or NRCS procedures that also generated 5 zones. Only in one of five years was another classification procedure (the NRCS Order 1 survey for 1993) better than the unsupervised classification at explaining yield variation.

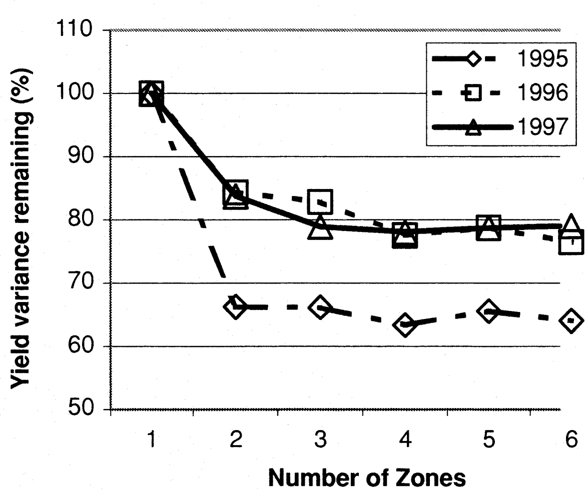

Figure 14. Portion of within–zone yield variance remaining after dividing Field 2 into management zones.

Table 4. Within–zone yield variance remaining after unsupervised classification (with optimum no. of zones), map overlay classification, and NRCS Order 1 classification.

Field	Classification Method	No. of Zones	Within Zone Yield Variance Remaining (%) 1993	1994	1995	1996	1997
1	Optimum unsupervised	5	88.9	71.1	90.2	69.6	69.0
	Map overlay	5	96.4	78.0	96.1	95.4	77.8
	NRCS Order 1 – 1991	3	98.4	73.4	97.7	99.4	90.2
	NRCS Order 1 – 1993	5	96.8	65.5	91.8	99.4	76.2
	NRCS Order 1 – 1997	7	97.6	74.7	89.2	97.4	85.9
2	Optimum unsupervised	4	n.a.	n.a.	63.3	77.5	78.0
	Map overlay	5	n.a.	n.a.	67.7	82.4	82.1

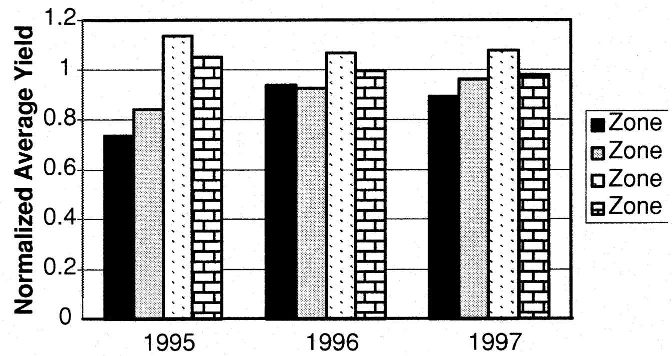

Figure 15. Calculated average normalized yield of four potential management zones defined for Field 2 (1995–1997).

Conclusions

Unsupervised classification of topographic attributes and soil EC for delineating potential within–field management zones was applied to two fields and the results analyzed. Principal component analysis of the input variables for one of the fields (Field 1) indicated that elevation and soil EC are the most important attributes to include when performing unsupervised classification in claypan soils. Slope and CTI are less important but may also be considered in the process. Evaluation of the resulting management zones using yield data indicated that, up to a point, within–zone yield variance generally decreases by increasing the number of zones. The case of little improvement in within–zone uniformity may suggest that either yield is uniform across the entire field, or important factors causing yield variability were not taken into consideration when determining the zones.

Yield data were used to determine the optimum number of zones when dividing a field. The optimum number of zones may vary from year to year and is mainly a function of weather and the crop planted. The number of zones decreases if adequate moisture conditions are present throughout the cropping season or if crops more tolerant to water stress are planted. Division of the field into a larger number of zones is generally recommended during years with below–average precipitation or when crops are subjected to water stress during critical periods of development.

Future work is required to investigate the weighting that should be given to each year's yield map in determining the optimum number of management zones. Frequency analysis of precipitation data could be a practical approach to determine the degree of importance of a given year. More importance or weight should be given to years with "typical" weather patterns. This analysis becomes more important when a limited number of yield maps is available and when the data include years of atypical weather.

Delineation of management zones based on topographic attributes and soil EC is a valid approach to capturing yield variability due to differences in plant water availability. Additional layers of information that may be considered important for characterizing the yield variability observed in a field, such as remote sensing images, crop scouting maps for diseases or insect damage, and soil fertility and pH maps, can easily be added to the unsupervised classification process. The methodology is fast, can be easily automated in commercially available GIS software, and has considerable advantages when compared to other methods for delineating within–field management zones.

Acknowledgements

The authors gratefully acknowledge Jon J. Fridgen and Scott T. Drummond for their help in preparing the data set used for this article.

References

Bell, J. C., C. A. Butler, and J. A. Thompson. 1995. Soil–terrain modeling for site–specific agricultural management. In *Site–specific Management for Agricultural Systems,* eds. P. C. Robert, R. H. Rust, and W. E. Larson, 209–227. Madison, Wisc.: ASA, CSSA, and SSSA.

Computer Aided Modeling. 1994. *Unscrambler User's Guide Ver. 5.5*. Trondheim, Norway: CAMO A/S.

Dunteman, G.H. 1989. *Principal Component Analysis*. Newbury Park, Calif.: Sage Publications.

ESRI. 1994. *Cell–based modeling with grid. Arc/Info Ver. 7.* Redlands, Calif.: Environmental Systems Research Institute.

Gallant, J. C., and J. P. Wilson. 1996. TAPES–G: A grid–based terrain analysis program for the environmental sciences. *Computers and Geosciences* 22(7): 713–722.

Irvin, B. J. 1996. Spatial information tools for delineating landform elements to support soil/landscape analysis. Ph.D. dissertation. University of Wisconsin, Madison.

Jaynes, D. B., T. S. Colvin, and J.Ambuel. 1995. Yield mapping by electromagnetic induction. In *Site–specific Management for Agricultural Systems,* eds. P. C. Robert, R. H. Rust, and W. E. Larson, 383–394. Madison, Wisc.: ASA, CSSA, and SSSA.

Jones, A. J., L. N. Mielke, C. A. Bartles, and C. A. Miller. 1989. Relationship of landscape position and properties to crop production. *J. Soil and Water Cons.* 44(4): 328–332.

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

_____. 1998. An evaluation of methods for determining site–specific management zones. In *Proc. North Central Extension—Industry Soil Fertility Conf.,* 133–139. Brookings, S. Dak.: Potash and Phosphate Institute.

McNeil, J. D. 1992. Rapid, accurate mapping of soil salinity by electromagnetic ground conductivity meters. In *Advances in Measurements of Soil Physical Properties: Bringing Theory Into Practice,* 201–229. SSSA Spec. Pub. 30. Madison, Wisc.: ASA, CSSA, and SSSA.

Mausbach, M. J., D. J. Lytle, and L. D. Spivey. 1993. Application of soil survey information to soil specific farming. In *Soil Specific Crop Management,* eds. P. C. Robert, R. H. Rust, and W. E. Larson, 57–68. Madison, Wisc.: ASA, CSSA, and SSSA.

Mulla, D. J., A. U. Bhatti, M. W. Hammond, and J. A. Benson. 1992. A comparison of winter wheat yield and quality under uniform versus spatially variable fertilizer management. *Agric. Ecosy. Environ.* 38: 301–311.

Sudduth, K. A., S. T. Drummond, S. J. Birrell, and N. R. Kitchen. 1996. Analysis of spatial factors influencing crop yield. In *Proc. 3rd Int'l. Conf. on Precision Agriculture,* eds. P. C. Robert, R. H. Rust, and W. E. Larson, 129–140. Madison, Wisc.: ASA, CSSA, and SSSA.

Sudduth, K. A., S. T. Drummond, S. J. Birrell, and N. R. Kitchen. 1997. Spatial modeling of crop yield using soil and topographic data. In *Proc. 1st European Conf. on Precision Agric., Vol. 1,* ed. J. V. Stafford, 439–447. Oxford, UK: Bios Scientific Publishers.

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 *Site–specific Management for Agricultural Systems,* eds. P. C. Robert, R. H. Rust, and W. E. Larson, 671–681. Madison, Wisc.: ASA, CSSA, and SSSA.

Tou, J. T., and R. C. Gonzalez. 1974. *Pattern Recognition Principles.* Reading, Mass.: Addison Wesley.

Wollenhaupt, N. C., D. J. Mulla, and C. A. Gotway Crawford. 1997. Soil sampling and interpolation techniques for mapping spatial variability of soil properties. In *The State of Site–specific Management for Agriculture,* eds. F. J. Pierce and E. J. Sadler, 19–53. Madison, Wisc.: ASA, CSSA, and SSSA.