# Management Zone Analyst (MZA): Software for Subfield Management Zone Delineation

Jon J. Fridgen, Newell R. Kitchen,* Kenneth A. Sudduth, Scott T. Drummond, William J. Wiebold, and Clyde W. Fraisse

## ABSTRACT

**Producers using site-specific crop management (SSCM) have a need for strategies to delineate areas within fields to which management can be tailored. These areas are often referred to as *management zones*. Quick and automated procedures are desirable for creating management zones and for testing the question of the number of zones to create. A software program called Management Zone Analyst (MZA) was developed using a fuzzy *c*-means unsupervised clustering algorithm that assigns field information into like classes, or potential management zones. An advantage of MZA over many other software programs is that it provides concurrent output for a range of cluster numbers so that the user can evaluate how many management zones should be used. Management Zone Analyst was developed using Microsoft Visual Basic 6.0 and operates on any computer with Microsoft Windows (95 or newer). Concepts and theory behind MZA are presented as are the sequential steps of the program. Management Zone Analyst calculates descriptive statistics, performs the unsupervised fuzzy classification procedure for a range of cluster numbers, and provides the user with two performance indices [fuzziness performance index (FPI) and normalized classification entropy (NCE)] to aid in deciding how many clusters are most appropriate for creating management zones. Example MZA output is provided for two Missouri claypan soil fields using soil electrical conductivity, slope, and elevation as clustering variables. Management Zone Analyst performance indices indicated that one field should be divided into either two (using NCE) or four (using FPI) management zones and the other field should be divided into four (using NCE or FPI) management zones.**

SITE-SPECIFIC CROP MANAGEMENT promotes the concept of identification and management of regions within the geographic area defined by field boundaries. Often referred to as management zones, these subfield regions typically represent areas of the field that are similar based on some quantitative measure(s) (e.g., topography, yield, and soil-test nutrients). Determination of subfield areas is difficult due to the complex combination of soil, biotic, and climate factors that may affect crop yield. These factors dynamically interact, further complicating the decisions of how to manage by zones. Three questions typically arise when considering managing by zones. One, what information should be used as a basis for creating different management zones within a field? Two, how can information be processed into unique management units (i.e., procedures for classification)? And three, how many unique zones should a field be divided into? Quick, efficient, and automated procedures are needed that address these questions.

A number of information sources have been used to delineate subfield management zones for SSCM. Traditional soil surveys often provide estimates of crop productivity for each soil map unit. In the USA, county soil surveys report the average yield of major crops and various soil properties by soil map unit; but the spatial scale of county soil surveys has often been found inadequate for use in SSCM (Mausbach et al., 1993). Digital elevation data collected using global positioning systems (GPS) or total station surveys have been used for classifying a field into management zones (McCann et al., 1996; Lark, 1998; MacMillan et al., 1998; van Alphen and Stoorvogel, 1998). Fleming et al. (2000) used aerial photographs of bare soil along with landscape position and the management experience of the producer to delineate within-field management zones. Because of the relationship of bulk soil apparent electrical conductivity ($EC_a$) to productivity on some soils (Kitchen et al., 1999, 2003), it has been used in the delineation of management units. Sudduth et al. (1996) and Fraisse et al. (2001a) used a combination of topographic attributes and soil $EC_a$ to delineate management zones. Long et al. (1994) investigated the accuracy and precision of field management maps created from several sources [e.g., soil survey map, aerial photograph, overlaying class values of kriged point data in a geographic information system (GIS)]. They concluded that aerial photographs of growing crops were the most accurate and precise for classifying a field into management units to predict grain yield. Imagery of a growing crop and yield data collected in the same year would be highly correlated and thus an accurate representation of crop production potential for that specific year (Boydell and McBratney, 1999).

Delineating zones based on topographic attributes and/or soil physical properties most often captures yield variability due to differences in plant available water and thus, crop production potential (McCann et al., 1996; van Alphen and Stoorvogel, 1998; Fraisse et al., 2001a).

J.J. Fridgen, ITD/Spectral Visions, 20407 South Neil Street Suite 2, Champaign, IL 61820; N.R. Kitchen, K.A. Sudduth, and S.T. Drummond, USDA–ARS, Cropping Syst. and Water Quality Res. Unit, Columbia, MO 65211; W.J. Wiebold, Dep. of Agron., University of Missouri, Columbia, MO 65211; and C.W. Fraisse, Agric. and Biol. Eng. Dep., Univ. of Florida, Gainesville, FL 32611. Received 22 Aug. 2002. *Corresponding author (kitchenn@missouri.edu).

Published in Agron. J. 96:100–108 (2004).


**Abbreviations:** $EC_a$, apparent soil electrical conductivity; FPI, fuzziness performance index; GIS, geographic information systems; ISODATA, Iterative Self-Organizing Data Analysis Technique; MZA, Management Zone Analyst; NCE, normalized classification entropy; SSCM, site-specific crop management.

The appropriate number of productivity zones was found to vary from year to year and was primarily affected by weather and the crop planted (Fraisse et al., 2001a). Fewer zones were required if sufficient moisture was received during the growing season or crops more tolerant to water stress were grown.

Although within-field management zones are often used to represent areas of equal production potential, they do have other uses. Fleming et al. (2000) suggested that producer-developed management zones based on fertility could be effective in creating variable-rate application maps. MacMillan et al. (1998) concluded that management zones are well suited for locating benchmark soil-sampling sites. Small, spatially coherent areas within fields may also be useful in relating yield to soil and topographic parameters for crop-modeling evaluation (Fraisse et al., 2001b).

In SSCM applications, cluster analysis procedures have been effectively used to identify regions of a field that are similar based on landscape attributes, fertility, or soil physical properties (Fraisse et al., 2001a). Stafford et al. (1998) used fuzzy clustering of combine yield monitor data to divide a field into potential management zones. Similarly, Boydell and McBratney (1999) divided a field into management zones using cotton (*Gossypium hirsutum* L.) yield estimates from satellite imagery. They concluded, however, that due to the effects of weather-induced variability on crop growth, at least 5 yr of yield data were required to create a stable number of management zones.

While a great deal of knowledge has been gained through analysis and interpretation of spatial data, procedures for generating management zones are not well prescribed for the end-user. Although many GIS packages contain the functions necessary for transforming spatial information into potential management zones, they can be cumbersome to use and require considerable time to learn. Producers have indicated a need for software to help them with decision-making and planning for variable within-field management (Wiebold et al., 1998). This paper reports on the development of software to assist researchers, consultants, and producers in creating management zones using quantitative soil, crop, and/or site information. The software, MZA, provides systematic procedures for delineating within-field management zones and for testing the results to evaluate the question of how many zones to use in a given field. Here we report on the theoretical basis of MZA, outline the steps and equations used, provide an example of MZA output for two fields, and list software specifications and availability.

## THEORETICAL BASIS FOR MANAGEMENT ZONE ANALYST DEVELOPMENT

Cluster analysis is the grouping of similar individuals into distinct classes called clusters. Literature reviews have shown many algorithm options for clustering data (Tou and Gonzalez, 1974; Hartigan, 1975), but there is no unified theory of clustering that is widely accepted (Milligan, 1996).

In supervised clustering, a combination of fieldwork, maps, aerial imagery, and personal experience are used to characterize specific sites that can then be used to represent the whole area (Mausel et al., 1990). These sites are referred to as training sites because the soil or vegetation reference information is used to train the classification algorithm for mapping the remainder of the data. After calculating multivariate statistical parameters for each training site, every data point both inside and outside the training sites is evaluated and assigned to the class for which it has the greatest likelihood of being a member (Jensen, 1996). We assumed the user may have no a priori knowledge of what information or areas should be used for training and therefore excluded supervised clustering as an option for MZA.

Unlike supervised clustering techniques, unsupervised clustering algorithms do not require the user to specify training areas. Unsupervised classification techniques produce natural groupings of the data in attribute space (Jensen, 1996; Irvin et al., 1997). Often, unsupervised classification is used to gain insight into the inherent structure of the data. The Iterative Self-Organizing Data Analysis Technique (ISODATA) (Tou and Gonzalez, 1974) is one of the more widely used unsupervised clustering algorithms. The ISODATA unsupervised classification algorithm calculates class means evenly distributed in the data space and then iteratively clusters the data points by minimizing the Euclidean distance from each data point to a class mean. Each iteration results in recalculation of class means and reclassification of data points with respect to the new means. This process continues until either the maximum number of iterations is reached or the number of data points in each class changes by less than the specified change threshold (ESRI, 1994; RSI, 1999). To effectively characterize output classes by mean vectors and a covariance matrix, ISODATA requires each variable in the data set to exhibit a roughly Gaussian distribution. Additionally, better results are obtained if all data exhibit similar variances (ESRI, 1994; Fraisse et al., 2001a; Irvin et al., 1997). These two requirements may require additional data set preparation before classification.

Unlike the ISODATA algorithm, the *c*-means (also known as *k*-means) algorithm does not require variables used in the classification to have similar variances or to follow a Gaussian distribution (Irvin et al., 1997). The *c*-means algorithm is based on the minimization of an objective function defined as the sum of squared distances from all data points in the cluster domain to the cluster center (i.e., the centroid). Similar to ISODATA, the *c*-means algorithm uses an iterative process to recalculate the cluster means and assign data points to clusters. The algorithm terminates when the specified convergence criterion (i.e., the amount of change in the cluster means) is met (Tou and Gonzalez, 1974).

Zadeh (1965) introduced the theory of fuzzy sets as a generalization of conventional set theory. Unlike conventional set theory, which allows an individual to belong to only one set, fuzzy set theory allows individuals to exhibit partial membership in each of a number of sets. Using this theory, Ruspini (1969) introduced the concept of fuzzy clustering, which allows any given data point to exhibit partial membership in a given class. The application of fuzzy set theory to clustering algorithms has allowed researchers to better account for the continuous variability in natural phenomena (Burrough, 1989).

One of the more extensively used clustering algorithms is the fuzzy *c*-means (also known as fuzzy *k*-means) algorithm. Fuzzy *c*-means uses a weighting exponent to control the degree to which membership sharing occurs between classes (Bezdek, 1981). Fuzzy *c*-means classification has been used to classify soil and landscape data (Burrough et al., 1992; McBratney and DeGruijter, 1992; Odeh et al., 1992; Irvin et al., 1997), yield data (Lark and Stafford, 1997; Lark, 1998; Stafford et al., 1998), and remotely sensed images (Ahn et al., 1999; Boydell and McBratney, 1999). Justification for using this algorithm when using soil information in the classification process has been documented (Odeh et al., 1992). We chose the fuzzy *c*-means algorithm for MZA, believing that many who would use the software would rely on continuum-based soil and landscape information as clustering inputs.

### Identifying a Measure of Similarity

Before a data cluster can be formed using the fuzzy *c*-means clustering procedure, it is necessary to establish an appropriate measure of similarity for assigning individual observations to a particular cluster. A measure of similarity is a procedure used to determine how similar an observation is to a cluster center. The measure of similarity most commonly used is a normalized distance from an observation to the cluster mean in attribute space (Tou and Gonzalez, 1974; Johnson, 1998). Thus, as the distance between the observation and cluster mean decreases, the similarity between the two increases. Euclidean distance, one of the more frequently used measures of similarity, gives equal weight to all measured variables and is sensitive to correlated variables (Bezdek, 1981). Geometrically, Euclidean distance generates clusters having a spherical shape, which in reality, rarely occurs in a soil system (Odeh et al., 1992). Johnson (1998) described a variant of the Euclidean distance method known as standardized Euclidean distance. This procedure computes the standard Euclidean distance between points using their standardized $Z$ scores.

The diagonal-distance method for measuring similarity is described by McBratney and Moore (1985) and Odeh et al. (1992). Like Euclidean distance, the diagonal method is sensitive to correlated variables. However, it does compensate for distortions in the assumed spherical shape of the clusters by weighting with the variances of the measured variables.

Another alternative is the Mahalanobis distance, which accounts for unequal variances as well as correlations between variables. It accomplishes this by including the pooled within-class variance–covariance matrix as an integral part of the distance calculation (Bezdek, 1981; McBratney and Moore, 1985; Odeh et al., 1992).

## MANAGEMENT ZONE ANALYST PROGRAM DESCRIPTION

Microsoft Visual Basic (Microsoft Corp., Redmond, WA) was used to develop MZA. Figure 1 depicts the decision structure and information flow through the software. Management Zone Analyst's function is the calculation of descriptive statistics, the delineation of management zones using the fuzzy *c*-means unsupervised classification algorithm, and the evaluation of the performance of the clustering by the number of clusters. To maximize utility with other software, MZA was designed to work with comma-delimited ASCII (the American Standard Code for Information Interchange) text files. Files of this format are easily exported from database, spreadsheet, graphics, or GIS software. The first line of the file must contain the variable names (i.e., column names), also separated by commas. There is no set maximum for the number of observations or variables in the data file. However, memory requirements increase as the number of observations and the number of variables increase. The analyses performed by the software have been tested with as many as 20 variables and 36 000 observations. Any quantitative data may be used in MZA. Sources of data may include, but are not limited to, soil properties (e.g., soil electrical conductivity), topographic attributes (e.g., elevation, slope, curvature), soil fertility, yield, and remotely sensed imagery.

### Calculation of Descriptive Statistics

Descriptive univariate and multivariate statistics of input variables are calculated before the clustering analysis is performed. Computations include the minimum and maximum values for each variable, the mean, standard deviation, coefficient of variation, between variable variance–covariance matrix, and correlation matrix. The variance–covariance matrix is provided for use in choosing the measure of similarity (explained in detail later).

### Fuzzy *c*-Means Background

The fuzzy *c*-means clustering algorithm was selected for the purpose of partitioning $n$ data observations in feature space into $c$ groups or clusters. "Fuzzy" refers to the shared membership between classes (Ruspini, 1969).

There are three primary matrices involved in the clustering process. First, there is the data we want to classify, the data matrix $\mathbf{Y}$, consisting of $n$ observations with $p$ classification variables each. Second is the cluster centroid matrix $\mathbf{V}$, consisting of $c$ cluster centroids located in the feature space defined by the $p$ classification variables. Finally, there is the fuzzy membership matrix $\mathbf{U}$, consisting of membership values to every cluster in $\mathbf{V}$ for each observation in $\mathbf{Y}$, bounded by the constraints for all $i = 1$ to $c$ and all $k = 1$ to $n$ that:

$$u_{ik} \in 0 - 1, \forall i, k \quad \text{and} \quad \sum_{i=1}^{c} u_{ik} = 1, \forall k \qquad [1]$$

The fuzzy *c*-means clustering algorithm attempts to locate minimal solutions to a selected objective function. The most commonly used objective function (and the

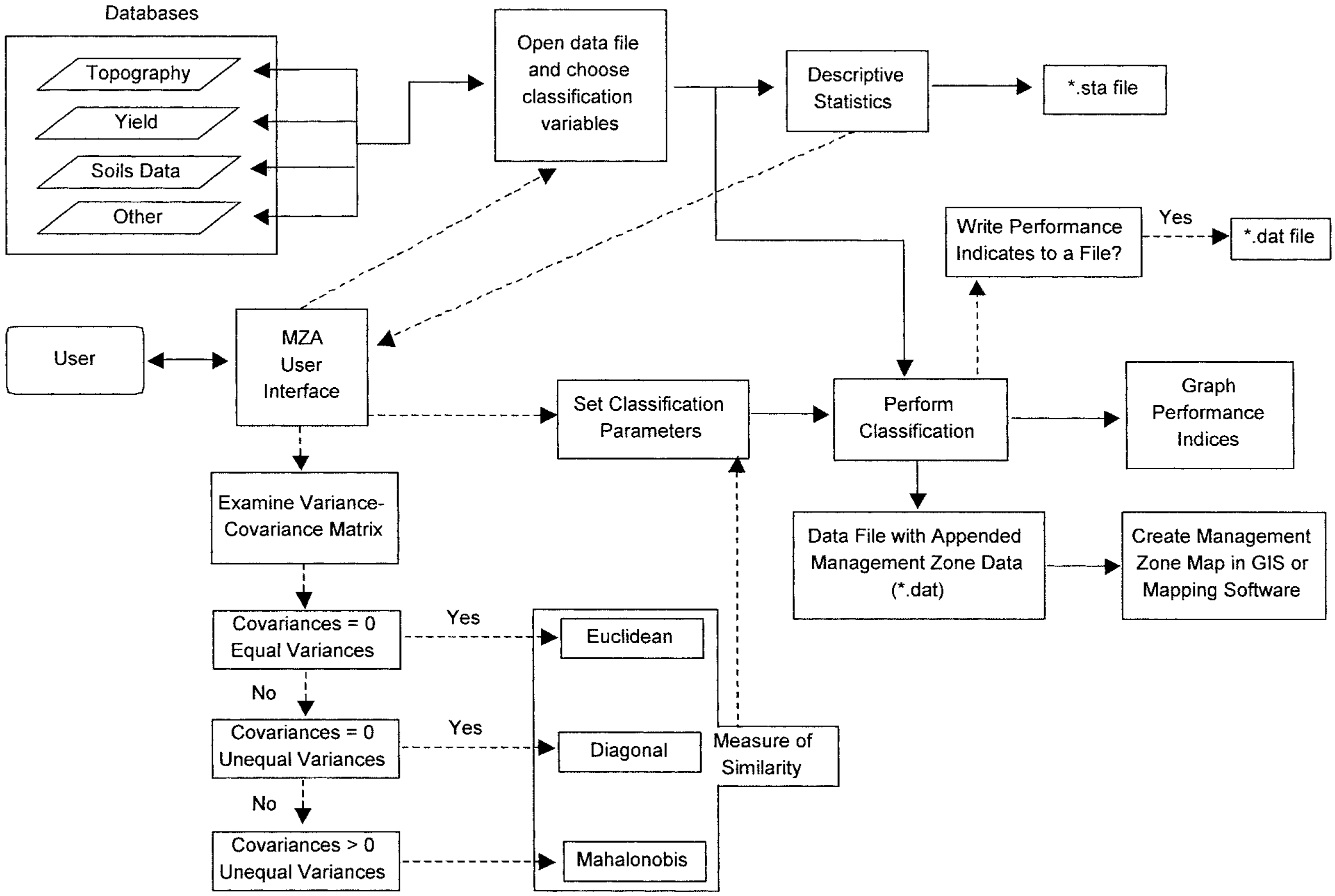


**Fig. 1. Diagram of information flow through Management Zone Analyst (MZA) software. Dashed lines represent user input.**

one used in MZA) is the weighted within-groups sum of squared errors objective function (Bezdek, 1981):

$$\mathbf{J}_m(\mathbf{U}, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m (d_{ik})^2 \qquad [2]$$

where $m$ = fuzziness exponent ($1 \leq m < \infty$) and $(d_{ik})^2$ = the squared distance in feature space between $y_k$ and $v_i$.

The fuzziness exponent ($m$) controls the amount of membership sharing that occurs between classes. As $m$ increases toward infinity, the amount of membership sharing increases, and the resulting classes become less distinct. Hard clusters (i.e., no membership sharing) occur as $m$ approaches a value of 1.

The distance in feature space between an observation in $\mathbf{Y}$ and a cluster centroid in $\mathbf{V}$ can be calculated in the following manner:

$$(d_{ik})^2 = \|y_k - v_i\|^2 = (y_k - v_i)' \, \mathbf{A}(y_k - v_i) \qquad [3]$$

where $y_k$ = the data observation $k$, consisting of $p$ classification variables; $v_i$ = the centroid of cluster $i$, consisting of $p$ classification variables; and $\mathbf{A}$ = positive definite, norm-inducing weight matrix of size $p \times p$.

The weight matrix $\mathbf{A}$ defines the distance-normalizing procedure. The result represents the distance between two points (vectors are used synonymously with points) in a linear vector space (Brogan, 1985). In the MZA software, the weight matrix may have three different forms depending on the covariance structure of the data. Management Zone Analyst provides the information necessary to select the proper measure of similarity in the descriptive statistics calculated before the delineation process.

The Euclidean distance should be used only for statistically independent variables exhibiting equal variances. In this case, $\mathbf{A}$ is the $p \times p$ identity matrix. This condition is rarely met in practice as even a change in the units of the classification variables will affect variances. A diagonal distance, whereby the identity matrix is adjusted by dividing each row of the matrix by the variance of the related classification variable, can address this problem. The diagonal distance is appropriate for statistically independent classification variables with unequal variances.

The third alternative is the Mahalanobis distance where $\mathbf{A}$ is defined as the inverse of the $p \times p$ sample variance–covariance matrix of $\mathbf{Y}$. The Mahalanobis distance, while slightly more computationally intensive, can account for situations where $\mathbf{Y}$ contains statistically dependent classification variables with unequal variances (Bezdek, 1981; Odeh et al., 1992). While all three distance methods are provided as options in MZA, a user relying on soil and landscape data will likely find the Mahalanobis option to be the most appropriate once the variance–covariance matrix is examined (Odeh et al., 1992).

## Cluster Performance Indices

While the iterative fuzzy $c$-means algorithm always converges to a local minimum of $\mathbf{J}_m$ starting from a given initial **U**, a different randomization of **U** might lead to a different local (or global) minimum (Xie and Beni; 1991; Bezdek, 1981; Ahn et al., 1999). To evaluate the characteristics of clustering by the number of clusters, two types of cluster validity functions were calculated on each fuzzy $c$-partition of **Y** produced by the fuzzy $c$-means clustering algorithm.

The fuzziness performance index (FPI) (Odeh et al., 1992; Boydell and McBratney, 1999) is a measure of the degree of separation (i.e., fuzziness) between fuzzy $c$-partitions of **Y** and is defined as:

$$\mathrm{FPI} = 1 - \frac{c}{(c-1)}\left[1 - \sum_{k=1}^{n}\sum_{i=1}^{c}\left(u_{ik}\right)^2 \Big/ n\right] \qquad [4]$$

Values of FPI may range from 0 to 1. Values approaching 0 indicate distinct classes with little membership sharing while values near 1 indicate nondistinct classes with a large degree of membership sharing.

Bezdek (1981) described a second measure of cluster validity known as the normalized classification entropy (NCE). The NCE models the amount of disorganization of a fuzzy $c$-partition of **Y** (Odeh et al., 1992; Lark and Stafford, 1997). The classification entropy ($H$) is defined by the function:

$$H(\mathbf{U};c) = -\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\log_a(u_{ik}) \Big/ n \qquad [5]$$

where logarithmic base $a$ is any positive integer. Values of $H$ will range from 0 to $\log_a(c)$. Bezdek (1981) reported that the endpoints of the range of $H$ do not accurately represent the amount of disorganization present (i.e., at $c = 1$, $H = 0$; at $c = n$, $H = 0$). To remedy this issue, he suggested the NCE:

$$\mathrm{NCE} = H(\mathbf{U};\mathbf{c})/[1 - (c/n)] \qquad [6]$$

The values of NCE will be similar to those of $H$ when $c$ is relatively small compared with $n$ [i.e., $(c/n)$ approaching 0]. However in situations where $(c/n)$ is large (i.e., approaching 1), NCE will produce substantially different results.

## Summary of Management Zone Analyst Steps

The algorithmic structure of the iterative fuzzy $c$-means algorithm (Bezdek, 1981) is

1. Choose the number of clusters $c$, with $2 \leq c < n$.
2. Choose the fuzziness exponent $m$, with $1 \leq m < \infty$.
3. Choose an appropriate measure of similarity for the distance metric $d_{ik}^2$.
4. Choose a value for the stopping criterion $\varepsilon$.
5. Choose a value for the maximum number of iterations $l_{\max}$.
6. Initialize $\mathbf{U}^0$ with random values meeting the specified constraints.
7. At iteration $l = 1, 2, 3, \ldots$, calculate updated $\mathbf{V}^l$ from $\mathbf{U}^{(l-1)}$, using:

$$v_i = \sum_{k=1}^{n}\left(u_{ik}\right)^m y_k \Big/ \sum_{k=1}^{n}\left(u_{ik}\right)^m,\ 1 \leq i \leq c \qquad [7]$$

8. Calculate updated $\mathbf{U}^l$ from updated $\mathbf{V}^l$, using:

$$u_{ik} = \left[\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}\right]^{-1} \qquad [8]$$

9. Stop when $l_{\max}$ is reached or when $\|\mathbf{U}^{(l)} - \mathbf{U}^{(l-1)}\| \leq \varepsilon$; otherwise go to Step 7.
10. Compute the cluster validity functions (FPI and NCE).

**Table 1. The header row and first 12 rows of an example input data set for Management Zone Analyst (MZA) file (with comma delineation removed).**

| Easting | Northing | Elevation | Soil EC† | Slope |
|---|---|---|---|---|
| 576000 | 4342200 | 264.8 | 71.5 | 0.56 |
| 576010 | 4342200 | 264.9 | 94.8 | 0.56 |
| 576020 | 4342200 | 264.9 | 98.1 | 0.27 |
| 576030 | 4342200 | 265.0 | 44.1 | 0.13 |
| 576040 | 4342200 | 265.0 | 42.1 | 0.11 |
| 576050 | 4342200 | 265.0 | 44.9 | 0.02 |
| 576060 | 4342200 | 265.0 | 43.5 | 0.1 |
| 576070 | 4342200 | 265.0 | 43.6 | 0.09 |
| 576080 | 4342200 | 265.0 | 41 | 0.18 |
| 576090 | 4342200 | 265.0 | 36.6 | 0.14 |
| 576100 | 4342200 | 264.9 | 36.6 | 0.02 |
| 576110 | 4342200 | 265.0 | 34.8 | 0.16 |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

† EC, electrical conductivity.

## MANAGEMENT ZONE ANALYST RESULTS FOR TWO CLAYPAN SOIL FIELDS

To illustrate MZA results, soil and landscape information for two Missouri claypan soil fields were used for creating potential management zones. Georeferenced measurements of soil $EC_a$, elevation, and slope (from elevation) were measured for a 36- and a 14-ha field (Field 1 and Field 2, respectively), kriged, and then gridded to a common 10-m cell as described for the Missouri field in Kitchen et al. (2003). This gave 3634 observations for Field 1 and 1308 observations for Field 2. Data files for each field were saved as a comma-delimited text file with the first row as labels for the columns, as shown in Table 1 (commas omitted for table). After importing a data file to MZA, the variables soil $EC_a$, elevation, and slope (Fig. 2 and 3) were selected as the clustering variables. These variables had previously been useful in delineating productivity zones for claypan soil fields (Fraisse et al., 2001a). MZA was then used to calculate descriptive statistics for the input variables. For both fields, the classification variables were found to have unequal variances and non-zero covariances; thus, the Mahalanobis measure of similarity option was chosen for the delineation procedure. Other option settings were fuzziness exponent = 1.5, maximum number of iterations = 300, convergence criterion = 0.0001, minimum number of zones = 2, and maximum number of zones = 8. When clustering using soils data, setting the fuzziness exponent between 1.2 and 1.5 will give reasonable results (Odeh et al., 1992). After the clustering process, the results were saved by MZA into a new text file as shown in Table 2. The output file is the same as the input file (Table 1) but

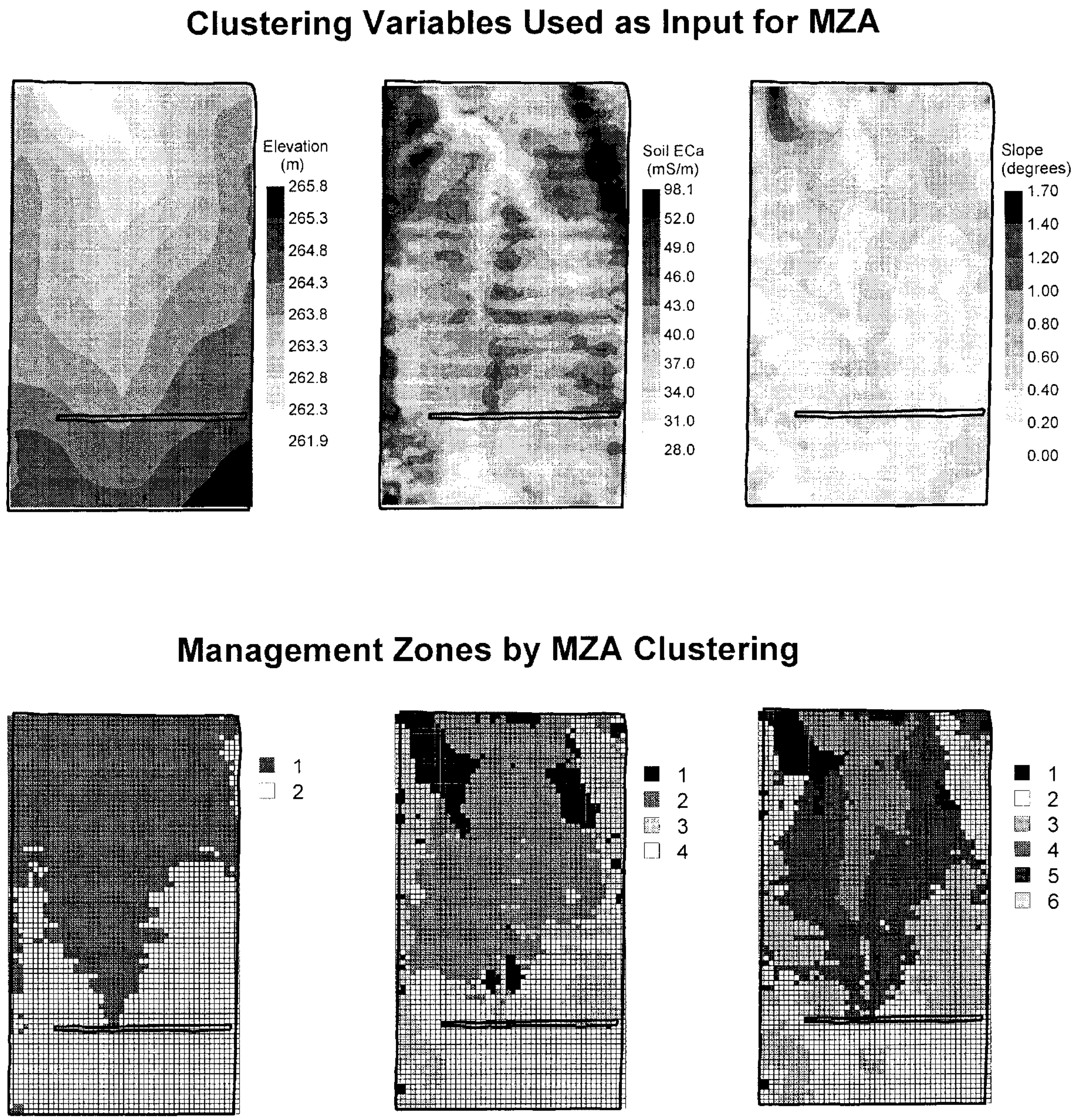


**Fig. 2. (Top) Apparent soil electrical conductivity ($EC_a$), elevation, and slope used as Management Zone Analyst (MZA) clustering variables for Field 1 and (bottom) MZA output for two, four, and six clusters.**

with appended columns for zone delineations from two zones to eight zones (shown as header labels of c2, c3, . . . c8). The exported data file from each field was imported into a mapping program and maps created for the two-, four-, and six-zone columns (bottom row of maps in Fig. 2 and 3).

The final step within MZA is a graphical representation of the FPI and NCE performance indices relative to cluster number to visually assess the optimal cluster number, similar to the approaches taken by others (Boydell and McBratney, 1999; Fraisse et al., 2001a). The optimal number of clusters for each computed index is when the index is at the minimum, representing the least membership sharing (FPI) or greatest amount of organization (NCE) as a result of the clustering process.

Management Zone Analyst also allows exporting the performance index data for graphing in other software. Results from the two indices were graphed for the two fields in Fig. 4. The minimum FPI was obtained for both fields with about four clusters. The minimum NCE was obtained with two clusters for Field 1 and with four clusters for Field 2. The final decision of how many clusters to use for creating management zones when the performance indices are dissimilar may require additional verification. For example, when developing productivity zones, verification of cluster number might be accomplished by comparing the within-zone yield variation as one increases the number of clusters (Fraisse et al., 2001a). Also, by comparing MZA output using different input variables, a user can assess which variables are most important for creating management zones.

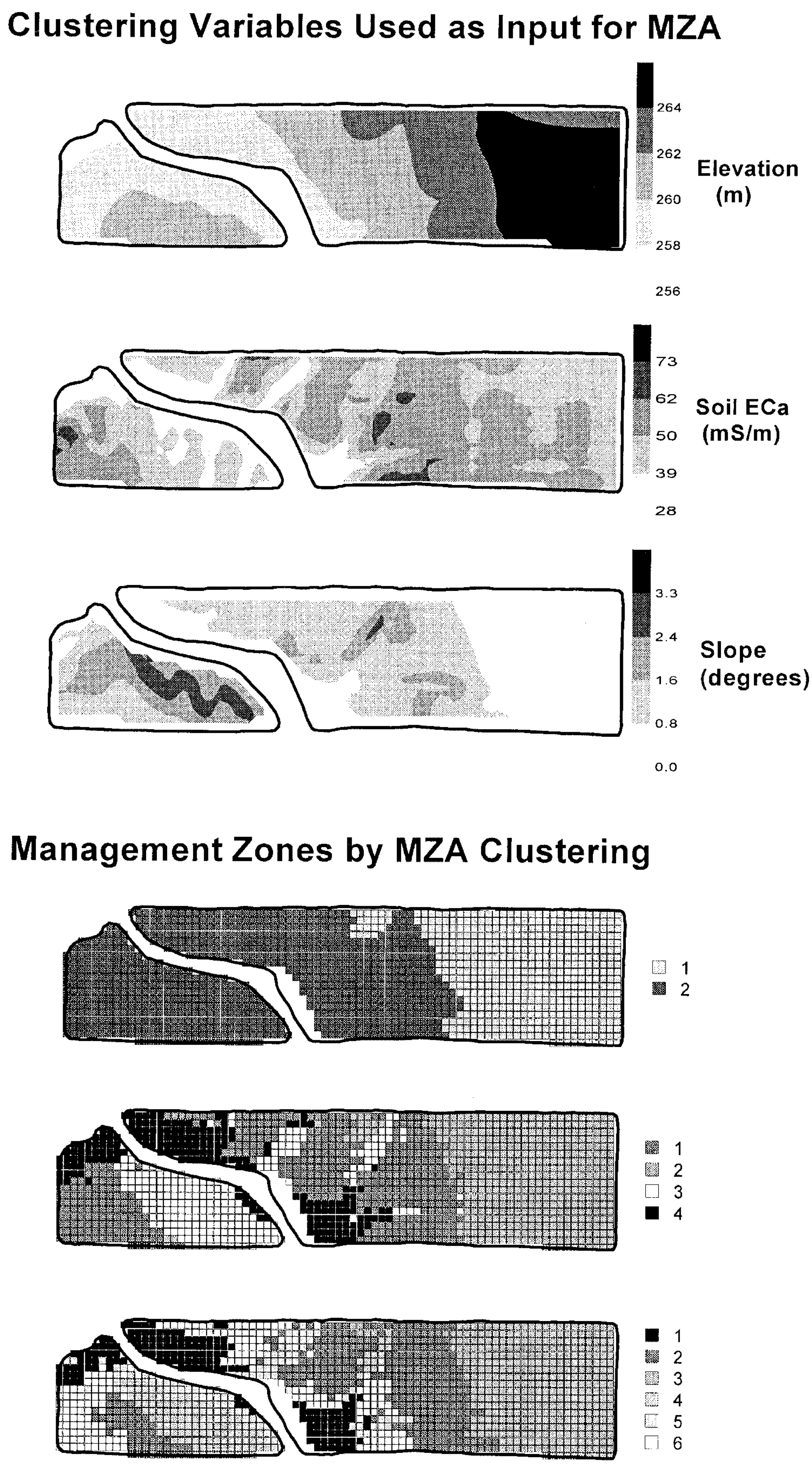


**Fig. 3. (Top) Apparent soil electrical conductivity ($EC_a$), elevation, and slope as MZA clustering variables for Field 2 and (bottom) MZA output for two, four, and six clusters.**

**Table 2. The header row and first 12 rows of data of an Management Zone Analyst (MZA) output file (with delineation removed). Columns 6 through 12 in this table are the classes that MZA created for the field from two to eight zones. The values in each of these columns represent the assigned class from the clustering procedure.**

| Easting | Northing | Elevation | Soil EC† | Slope | c2 | c3 | c4 | c5 | c6 | c7 | c8 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 576000 | 4342200 | 264.8 | 71.5 | 0.56 | 1 | 2 | 3 | 2 | 6 | 2 | 7 |
| 576010 | 4342200 | 264.9 | 94.8 | 0.56 | 1 | 2 | 3 | 2 | 6 | 2 | 7 |
| 576020 | 4342200 | 264.9 | 98.1 | 0.27 | 1 | 2 | 3 | 2 | 6 | 2 | 7 |
| 576030 | 4342200 | 265.0 | 44.1 | 0.13 | 2 | 1 | 4 | 4 | 3 | 1 | 1 |
| 576040 | 4342200 | 265.0 | 42.1 | 0.11 | 2 | 1 | 4 | 4 | 3 | 1 | 1 |
| 576050 | 4342200 | 265.0 | 44.9 | 0.02 | 2 | 1 | 4 | 4 | 3 | 1 | 1 |
| 576060 | 4342200 | 265.0 | 43.5 | 0.1 | 2 | 1 | 4 | 4 | 3 | 1 | 1 |
| 576070 | 4342200 | 265.0 | 43.6 | 0.09 | 2 | 1 | 4 | 4 | 3 | 1 | 1 |
| 576080 | 4342200 | 265.0 | 41 | 0.18 | 2 | 1 | 4 | 4 | 3 | 1 | 1 |
| 576090 | 4342200 | 265.0 | 36.6 | 0.14 | 2 | 1 | 4 | 5 | 2 | 5 | 8 |
| 576100 | 4342200 | 264.9 | 36.6 | 0.02 | 2 | 1 | 4 | 5 | 2 | 5 | 8 |
| 576110 | 4342200 | 265.0 | 34.8 | 0.16 | 2 | 1 | 4 | 5 | 2 | 5 | 8 |
| . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . |
| . | . | . | . | . | . | . | . | . | . | . | . |

† EC, electrical conductivity.

As previously indicated, multiple outcomes are inherent to the MZA clustering algorithm. In one study, as few as one outcome (at four clusters) and as many as seven outcomes (at eight clusters) were possible using the fuzzy *c*-means algorithm (Kitchen et al., 2002). In MZA, initial cluster membership of observations is randomly assigned, and convergence to a local or global minima may vary depending on that starting point (Xie and Beni, 1991; Bezdek, 1981; Ahn et al., 1999). We found, when using the same three delineation variables as in the example above, that multiple outcomes were most different when the number of clusters was three or fewer (Kitchen et al., 2002). Each data set will behave uniquely, and we encourage users to test each data set by running MZA and then evaluating the results using the performance indices and other validation methods as outlined by Kitchen et al. (2002).

## MANAGEMENT ZONE ANALYST SPECIFICATIONS AND AVAILABILITY

Management Zone Analyst 1.0 was developed using Microsoft Visual Basic 6.0 and operates on any computer running Microsoft Windows (95 or newer). The MZA executable and associated components require approximately five megabytes of hard disk space and a minimum of eight megabytes of random access memory (RAM). As the number of clusters and/or the size of the data set increases, memory requirements also increase. Due to the large number of computations performed during the zone delineation process, processor speed will influence the time required to create the specified number of clusters. The latest version of MZA is available free from the Internet at http://www.fse.missouri.edu/ars/decision_aids.htm (verified 30 Oct. 2003). Included with the software download is a user guide that illustrates the steps for running the program. Management Zone Analyst program code is available in Fridgen (2000).

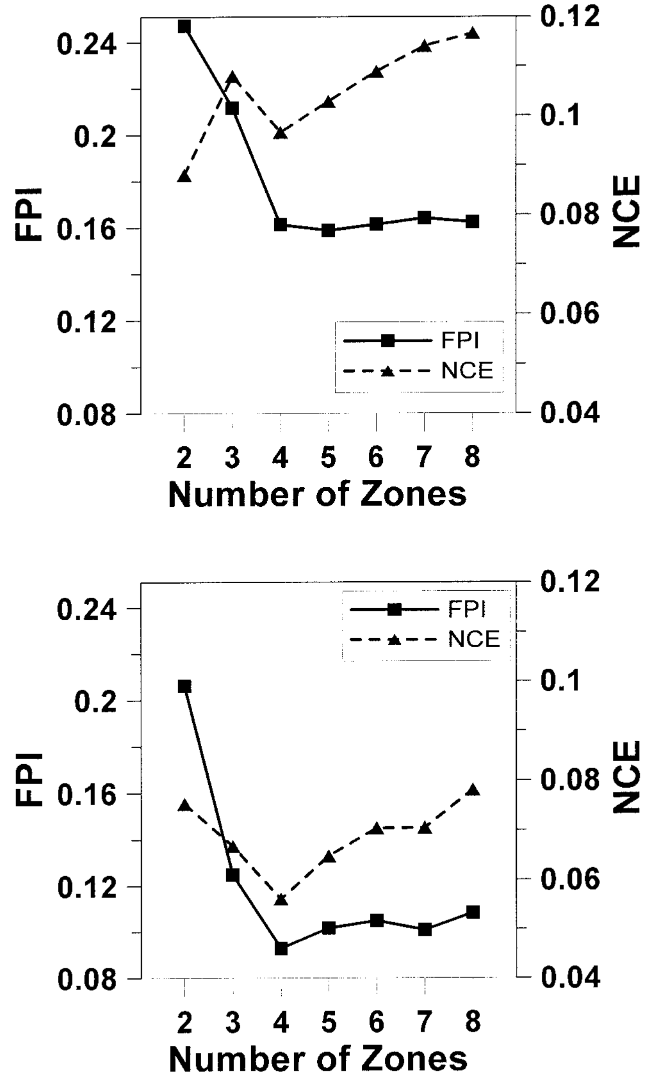


**Fig. 4. Fuzziness performance index (FPI) and normalized classification entropy (NCE) as calculated by MZA for (top) Field 1 and (bottom) Field 2. Generally, the best classification occurs when membership sharing (FPI) and/or the amount of class disorganization (NCE) is at a minimum with the least number of classes used.**

## ACKNOWLEDGMENTS

We thank the following for financial support for this project: North Central Soybean Research Program, United Soybean Board, the Foundation for Agronomic Research (FAR), and the USDA-CSREES National Research Initiative and Special Water Quality Grants program. We also thank M. Krumpelman, B. Mahurin, and M. Volkmann for their assistance in collecting field and crop measurements.

## REFERENCES

Ahn, C.W., M.F. Baumgardner, and L.L. Biehl. 1999. Delineation of soil variability using geostatistics and fuzzy clustering analyses of hyperspectral data. Soil Sci. Soc. Am. J. 63:142–150.

Bezdek, J.C. 1981. Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York.

Boydell, B., and A.B. McBratney. 1999. Identifying potential within-field management zones from cotton yield estimates. p. 331–341. *In* J.V. Stafford (ed.) Precision agriculture '99. Proc. European Conf. on Precision Agric., 2nd, Odense Congress Cent., Denmark. 11–15 July 1999. SCI, London.

Brogan, W.L. 1985. Modern control theory. Prentice Hall, Englewood Cliffs, NJ.
Burrough, P.A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. J. Soil Sci. 40:477–492.
Burrough, P.A., R.A. Macmillan, and W. VanDeursen. 1992. Fuzzy classification methods for determining land suitability from soil profile observations and topography. J. Soil Sci. 43:193–210.
[ERSI] Environmental Systems Research Institute. 1994. Grid commands. ERSI, Redlands, CA.
Fleming, K.L., D.G. Westfall, D.W. Wiens, and M.C. Brodah. 2000. Evaluating farmer developed management zone maps for variable rate fertilizer application. Precis. Agric. 2:201–215.
Fraisse, C.W., K.A. Sudduth, and N.R. Kitchen. 2001a. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. Trans. ASAE 44(1):155–166.
Fraisse, C.W., K.A. Sudduth, and N.R. Kitchen. 2001b. Calibration of the Ceres-Maize model for simulating site-specific crop development and yield on claypan soils. Appl. Eng. Agric. 17(4):547–556.
Fridgen, J.J. 2000. Development and evaluation of unsupervised clustering software for sub-field delineation of agricultural fields. M.S. thesis. Univ. of Missouri, Columbia.
Hartigan, J.A. 1975. Clustering algorithms. John Wiley & Sons, New York.
Irvin, B.J., S.J. Ventura, and B.K. Slater. 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. Geoderma 77:137–154.
Jensen, J.R. 1996. Introductory digital image processing: A remote sensing perspective. Prentice Hall, Upper Saddle River, NJ.
Johnson, D.E. 1998. Applied multivariate methods for data analysis. Brooks/Cole Publ., Pacific Grove, CA.
Kitchen, N.R., S.T Drummond, E.D. Lund, K.A. Sudduth, and G.W. Buchleiter. 2003. Soil electrical conductivity and topography related to yield for three contrasting soil–crop systems. Agron. J. 95: 483–495.
Kitchen, N.R., J.J. Fridgen, K.A. Sudduth, S.T. Drummond, W.J. Wiebold, and C.W. Fraisse. 2002. Procedures for evaluating unsupervised classification to derive management zones. *In* P.C. Robert et al. (ed.) Precision agriculture [CD-ROM]. Proc. Int. Conf., 6th, Minneapolis, MN. 14–17 July 2002. ASA, CSSA, and SSSA, Madison, WI.
Kitchen, N.R., K.A. Sudduth, and S.T. Drummond. 1999. Soil electrical conductivity as a crop productivity measure for claypan soils. J. Prod. Agric. 12:607–617.
Lark, R.M. 1998. Forming spatially coherent regions by classification of multivariate data: An example from the analysis of maps of crop yield. Int. J. Geogr. Inf. Sci. 12:83–98.
Lark, R.M., and J.V. Stafford. 1997. Classification as a first step in the interpretation of temporal and spatial variation of crop yield. Ann. Appl. Biol. 130:111–121.
Long, D.S., G.R. Carlson, and S.D. DeGloria. 1994. Quality of field management maps. p. 251–271. *In* P.C. Robert, R.H. Rust, and W.E. Larson (ed.) Site-specific management for agricultural systems. Proc. Int. Conf., 2nd, Minneapolis, MN. 27–30 Mar. 1994. ASA, CSSA, and SSSA, Madison, WI.
MacMillan, R.A., W.W. Pettapiece, L.D. Watson, and T.W. Goddard. 1998. A landform segmentation model for precision farming. p. 1335–1346. *In* P.C. Robert et al. (ed.) Precision agriculture. Proc. Int. Conf., 4th, St. Paul, MN. 19–22 July 1998. ASA, CSSA, and SSSA, Madison, WI.
Mausbach, M.J., D.J. Lytle, and L.D. Spivey. 1993. Application of soil survey information to soil specific farming. p. 57–68. *In* P.C. Robert et al. (ed.) Soil specific crop management. Proc. Int. Conf., Minneapolis, MN. 14–16 Apr. 1992. ASA, CSSA, and SSSA, Madison, WI.
Mausel, P.W., W.J. Kamber, and J.K. Lee. 1990. Optimum band selection for supervised classification of multispectral data. Photogramm. Eng. Remote Sens. 56:55–60.
McBratney, A.B., and J.J. DeGruijter. 1992. A continuum approach to soil classification by modified fuzzy k-means with extragrades. J. Soil Sci. 43:159–175.
McBratney, A.B., and A.W. Moore. 1985. Application of fuzzy sets to climatic classification. Agric. For. Meteorol. 35:165–185.
McCann, B.L., D.J. Pennock, C. van Kessel, and F.L. Walley. 1996. The development of management units for site specific farming. p. 295–302. *In* P.C. Robert, R.H. Rust, and W.E. Larson (ed.) Precision agriculture. Proc. Int. Conf., 3rd, Minneapolis, MN. 23–26 June 1996. ASA, CSSA, and SSSA, Madison, WI.
Milligan, G.W. 1996. Clustering validation: Results and implications for applied analyses. p. 341–375. *In* P. Arabie, L.J. Hubert, and G. De Soete (ed.) Clustering and classification. World Sci. Publ., River Edge, NJ.
Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1992. Soil pattern recognition with fuzzy-c-means: Application to classification and soil–landform interrelationships. Soil Sci. Soc. Am. J. 56: 505–516.
[RSI] Research Systems Incorporated. 1999. ENVI user's guide. Version 3.2. RSI, Boulder, CO.
Ruspini, E.H. 1969. A new approach to clustering. Inf. Control. 15: 22–32.
Stafford, J.V., R.M. Lark, and H.C. Bolam. 1998. Using yield maps to regionalize fields into potential management units. p. 225–237. *In* P.C. Robert et al. (ed.) Precision agriculture. Proc. Int. Conf., 4th, St. Paul, MN. 19–22 July 1998. ASA, CSSA, and SSSA, Madison, WI.
Sudduth, K.A., S.T. Drummond, S.J. Birrell, and N.R. Kitchen. 1996. Analysis of spatial factors influencing crop yield. p. 129–140. *In* P.C. Robert, R.H. Rust, and W.E. Larson (ed.) Precision agriculture. Proc. Int. Conf., 3rd, Minneapolis, MN. 23–26 June 1996. ASA, CSSA, and SSSA, Madison, WI.
Tou, J.T., and R.C. Gonzalez. 1974. Pattern recognition principles. Addison-Wesley, Reading, MA.
van Alphen, B.J., and J.J. Stoorvogel. 1998. A methodology to define management units in support of an integrated, model-based approach to precision agriculture. p. 1267–1278. *In* P.C. Robert et al. (ed.) Precision agriculture. Proc. Int. Conf., 4th, St. Paul, MN. 19–22 July 1998. ASA, CSSA, and SSSA, Madison, WI.
Wiebold, W.J., K.A. Sudduth, J.G. Davis, D.K. Shannon, and N.R. Kitchen. 1998. Determining barriers to adoption and research needs of precision agriculture [Online]. Report to the North Central Soybean Research Program. Available at http://www.fse.missouri.edu/mpac/pubs/parpt.pdf (verified 30 Oct. 2003). Missouri Precision Agric. Cent., Univ. of Missouri, Columbia.
Xie, X.L., and G. Beni. 1991. A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13(8):841–847.
Zadeh, L.A. 1965. Fuzzy sets. Inf. Control. 8:338–353.