Published online June 17, 2005

Defining Yield-Based Management Zones for Corn–Soybean Rotations

A. Brock, S. M. Brouder,* G. Blumhoff, and B. S. Hofmann

ABSTRACT

Unsupervised clustering has been proposed for developing georeferenced agronomic information into management zones (MZs). Our objectives were to use fuzzy c-means clustering to identify yield-based MZs, and to compare spatial association and agreement among corn (*Zea mays* L.) yield-based (CYB) MZs, soybean [*Glycine max* (L.) Merr.] yield-based (SYB) MZs, and published soil survey map units. Six years of yield monitor data (three per species) from four fields were used with the clustering software MZ Analyst. Clustering success was evaluated with four performance measures. Two measures of variance reduction and the fuzziness performance index (FPI) indicated clustering optimization with 4 to 6 MZs. In contrast, the normalized classification entropy (NCE) indicated that yield data were optimally organized with only 2 MZs. On average, the 4-MZ delineation reduced the yield variance to 40% of the whole field variance (corn within CYB MZs and soybean within SYB MZs); mean relative yields within MZs were significantly different from each other, ranging from 23% below to 12% above the whole-field mean. With 4 MZs, CYB and SYB MZs were significantly associated in all fields, but weighted agreement between CYB and SYB MZs was only slight ($0.06 \leq K_w \leq 0.34$), indicating crop-specificity in MZ delineation. In general, highest yielding MZs were significantly associated with areas mapped as a poorly drained, level soil series while lower yielding MZs corresponded to map units for eroded or more slopping soils. However, clustering yields by soil series reduced yield variance less than unsupervised, yield-based clustering. Routine application of MZ Analyst likely requires more decision support for identifying clustering success.

A BASIC PREMISE OF precision agriculture or site-specific management is that management can be implemented on a spatial scale smaller than that of a whole field. A major barrier to widespread adoption remains the lack of simple decision rules and protocols on how to practically and routinely delineate such management zones (MZs). Since the beginning of the precision agriculture technology era, patterns of yield variability have been considered important for variable rate nutrient management because productivity expectations or yield goals influence rate recommendations in much of the Midwest (Lamb et al., 1997). Many university recommendations for fertilizer management are based on identifying soil productivity potential and providing fertilizer to compliment the soils native nutrient supply in meeting the needs of a high yielding crop. Theoretically, consistent, within-field variation in yield should reflect within-field variation in soil productivity potential and related, soil-specific input needs. Widespread availability of combine-mounted yield monitors has made the collection of information on spatial variability in yield practically possible but the utility of this information for MZ delineation remains largely unexplored.

A decade ago, little was known about the spatiotemporal variability in corn and soybean yields (Jaynes and Colvin, 1997). Numerous, recent studies have used correlation and regression to examine temporal stability of spatial yield patterns within fields, and a common result has been the lack of stable yield patterns over time. For example, Lamb et al. (1997) found that corn yields observed in 1 yr accounted for between 4 and 42% of the variation in yields observed in subsequent years. Likewise, Jaynes and Colvin (1997) reported rank correlation coefficients of –0.09 to 0.54 and 0.03 to 0.52 for within field comparison of multiyear corn and soybean data, respectively. The lack of strong temporal correlation among yield data sets collected from a given field has been cited as evidence that management based on site-specific prediction of yield may not be successful. However, a MZs approach requires that we identify areas within a field that behave similarly through time. Jaynes et al. (2003) suggest that such MZs may be identifiable and useful for management even when fields exhibit poor temporal stability in relative yield patterns; they advocate cluster analysis instead of correlation and regression for converting long-term yield data into management information. Unsupervised clustering algorithms have been proposed for delineating MZs from yield monitor data (Lark and Stafford, 1997 and Stafford et al., 1998). This multivariate clustering approach groups similar observations into distinct classes but does not require the user to direct or train the classification algorithm with benchmark data, a requirement for supervised clustering. Because they do not require any a priori training, unsupervised clustering techniques are simpler to apply than supervised techniques and, therefore, may be better suited to general, on-farm applications (Fridgen et al., 2004).

In the eastern cornbelt of the USA the dominant production system is corn grown in annual rotation with soybean, and recommendations, especially for fertility management, are often for the whole rotation and not for the crop within the rotation. Temporal variation in rainfall pattern coupled with the varying drainage potential of the major agricultural soils are a governing factor for year-to-year variation in whole-field yields as well as the within-field spatial variation in yield productivity potential. Numerous reports have documented the extent to which both corn and soybean yields can be effected by inadequate drainage (reviewed by Evans and Fausey, 1999); where needed, artificial drainage enhance-

Department of Agronomy, 3351 Lilly Hall of Life Sciences, 915 W. State St., Purdue Univ., West Lafayette, IN 47907-2054. Received 21 Aug. 2004. *Corresponding author (sbrouder@purdue.edu).

Published in Agron. J. 97:1115–1128 (2005).
Site-Specific Management
doi:10.2134/agronj2004.0220

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: ANOVA, analysis of variance; CYB, corn yield-based; FPI, fuzziness performance index; MZs, management zones; NCE, normalized classification entropy; SYB, soybean yield-based; YB, yield-based.

ments including subsurface tiles are expected to improve the yields of both crop species. A logical inference drawn from this observation is that both crops will experience some degree of common, site- or soil-specific productivity limitations on fields with pronounced in-field variation in drainage potential. To date, however, few studies have addressed agreement among spatiotemporal yield patterns of multiple crop species grown in rotation systems. Consequently, the agreement among MZs derived from cluster analysis of crop-specific yield patterns has not been quantitatively examined.

Finally, despite the proliferation of software to record and manipulate spatial data, farm managers still lack user-friendly tools for converting farm data into MZs. The recently developed software package, MZ Analyst (Fridgen et al., 2004), uses the fuzzy c-means clustering algorithm to delineate MZs and it is intended for application to commonly collected farm management datalayers. A desirable attribute of the fuzzy c-means algorithm for soil, landscape, and crop physiological characteristics data is membership sharing between classes, a clustering feature appropriate for continuous parameters (Burrough, 1989). The MZ Analyst software is suitable for but has yet to be applied to the delineation of MZs from multiple years of yield monitor data.

The overall goal of this study was to advance our understanding of both theoretical and practical aspects of MZ delineation in nonirrigated, corn–soybean production systems. The principle objectives were (i) to determine whether unsupervised cluster analysis applied to yield monitor data could delineate potential MZs with unique productivities, (ii) to characterize the spatial agreement among MZs delineated by crop-specific (corn or soybean) and combined crop (corn plus soybean) productivities, and (iii) to characterize the spatial agreement between productivity based MZs and soil map units as described by county soil surveys. A secondary objective was to evaluate the prototype unsupervised clustering software, MZ Analyst.

MATERIALS AND METHODS

This study was conducted between 1996 and 2002 at the Davis Purdue University Agricultural Center in east-central Indiana (40°25′ N lat; 85°15′ W long). The four study fields, D, F, R, and V (11.3, 5.8, 13.2, and 15 ha, respectively), were located within 3 km of each other. The fields have histories of rotational corn–soybean production with a combination of no-till and conservation tillage. All fields were planted between 22 April and 12 June with seeding rates of either 64 250 to 77 000 corn seeds ha^{-1} or 519 000 to 544 000 soybean seeds ha^{-1}; corn crops were planted in 0.76-m rows while soybean crops were planted in 0.19-m rows. University recommendations were used to guide fertilizer applications (Vitosh et al., 1996) and pest management measures were implemented as needed in accordance with an integrated pest management approach.

Three years each of corn and soybean yield monitor data were collected for all study fields except D where only 2 yr of data were available for each crop. Yield data were collected at 1-s intervals using yield-monitoring systems with pressure based sensors (AgLeader 2000 or PF3000, Ag Leader Technology, Ames, IA) coupled to a differentially corrected global positioning system. Raw yield monitor data were adjusted to standard moisture content (15 and 13.5% for corn and soybean, respectively), analyzed for time lags on a per field and year basis and rectified. Yield data were analyzed for mild and severe outliers based on numeric difference from the interquartile range of a given dataset (as described by Devore, 2000). All severe outliers were removed, while mild outliers were individually evaluated and removed if they appeared to be associated with an operator artifact (e.g., actual harvest swath $<$ header width).

Values along a harvest pass were smoothed with a 3-point running average; the population of smoothed values from a given field was analyzed to identify values $\geq$95th percentile. The mean of this subpopulation of the data set was identified as the field-year maximum yield and relative yields were calculated as a percentage of this maximum. Relative yield values were interpolated to a 5-m grid using inverse distance weighting (Spatial Analyst extension, ArcView 3.2; ESRI, 1999). To reduce speckling and increase the likelihood of identifying classes with greater continuity, a 3 by 3 low-pass, standard mean convolution filter was applied to each relative yield map (Smooth Surface function in Grid Analyst extension [ESRI, 1999]). Mean relative corn, soybean, and combined corn–soybean yields were calculated for the 5-m grid maps. For each field, Pearson correlation was used to evaluate the relationships among individual years and between individual years and the mean relative yields.

Fields were classified into regions or MZs of similarity by applying fuzzy-c means unsupervised clustering to relative yield data (MZ Analyst; Fridgen, 2000). For a given field, input data for clustering were either all years of monitor data for one crop species resulting in the development of corn or soybean yield-based (CYB and SYB, respectively) MZs or all years of data irrespective of crop species resulting in general yield-based (YB) MZs. To determine the most appropriate measure of similarity (Euclidean, Diagonal, or Mahalanobis distances), input data layers were evaluated for equality of variance and statistical dependence using Levine's test and Pearson correlations. A fuzziness exponent of 1.5 was chosen following Lark and Stafford (1997) and Fridgen et al. (2004). For each field, we examined between two and six zones (convergence criterion of 0.0001 with $\leq$500 iterations).

Clustering success was examined with four performance measures. The FPI and the NCE were calculated according to Fridgen et al. (2004) using MZ Analyst. Two measures of variance reduction resulting from clustering were also evaluated. For each field-year, the reduction of total within-MZ yield variance following clustering into a given number of MZs as compared with yield variance observed on a whole-field basis was calculated following Fraisse et al. (2001) as follows:

$$\% \text{ of total variance} = 100 \times (S_T^2 \div S_{WF}^2) \quad [1]$$

where S_{WF}^2 is the variance calculated from all observations for a given field-year and S_T^2 is the summed, weighted variances from the 1 to c individual MZs being examined or

$$S_T^2 = S_1^2 + S_2^2 + S_3^2 + \ldots + S_c^2 \quad [2]$$

The weighted variance in MZ c was calculated as

$$S_c^2 = \left(\frac{1}{n_z}\sum_{i=1}^{n_z}(Y_i - \overline{Y}_c)^2\right) \times \left(\frac{n_z}{n_T}\right) \quad [3]$$

where n_z is the number of observations in MZ c, n_T is the total number of observations in the map, Y_i is the measured yield of the ith observation in MZ c, and $\overline{Y}_c$ is the mean of all i yield observations in MZ c. The percent above the minimum

observed variance, a second measure of variance reduction with clustering, was calculated as follows:

$$\% \text{ above minimum variance} = 100 \times [(S^2_{T-x} - S^2_{T-min}) \div S^2_{T-min}] \quad [4]$$

where S^2_{T-x} is the S^2_T when x MZs are considered and S^2_{T-min} is the lowest S^2_T among all the clusterings considered (e.g., whole field or one MZ to a maximum six MZs) (Fridgen, 2000).

For each classification strategy (CYB, SYB, or YB MZs), number of MZs, crop species, and field were used as class variables and percentage of total variance or percentage above minimum variance were the response variables in an analysis of variance (ANOVA) to identify significant main factors and interactions (GLM; SAS Inst., 2000). Nonsignificant main effects and interactions ($p > 0.25$) were pooled into the error terms. Fisher's LSD comparisons were used to identify the classification with lowest number of MZs where response variables did not differ significantly from those resulting from the maximum order clustering of six MZs.

Following derivation of the optimum number of CYB, SYB, and YB MZs based on evaluation of NCE, FPI and variance reduction, ANOVA was also used to determine whether significant differences existed in yields among MZs. For each field, MZ identity was used as a class variable and all individual years of yield data as well as mean relative yield data were used as response variables. A t test equivalent to Fisher's LSD comparisons for equal sample sizes was used to identify significant differences in response variables among MZs.

Degree of association among the three classification strategies (CYB, SYB, and YB) was assessed for each clustering order (FREQ; SAS Inst., 2000). Within a given field, MZs were ranked from highest (value = 1) to lowest yielding (value = 2–6). Cramer's V, derived from the Pearson chi-square, was used to characterize general association between two strategies (e.g., SYB vs. CYB) for a given clustering order. A Mantel-Haenszel procedure with standardized midranks (modified ridit scrores) was used to detect linear correlation between two classification strategies as the alternative to the null hypothesis of no association between classification strategies in assignment of observations to yield level clusters. To evaluate agreement among CYB, SYB, and YB MZs for each clustering order, the proportion of overall raw agreement was calculated as follows:

$$P_O = \frac{1}{N}\sum_{i=1}^{C} n_{ii} \quad [5]$$

where N is the sample size, C is the clustering order or table dimension, and n_{ii} is the frequency of a given cell in the main diagonal of an n_{ij} table. Weighted agreement among table cells was evaluated with Kappa statistics. The simple Kappa coefficient, K_S was estimated as follows:

$$K_S = \frac{P_O - P_E}{1 - P_E} \quad [6]$$

where P_E is the equivalent proportion for the table of the expected values for diagonal cells under the null hypothesis of random association. A weighted Kappa coefficient, K_W, was calculated as described by Stokes et al. (2000) using Ciccetti-Allison weights to permit more weight to be given to agreement in cells closest to the diagonal in square contingency tables and less weight to be given to cells further from the diagonal.

As described above for determining yield differences among zones, ANOVA was used to evaluate yield differences among soil types. For each field, soil series was used as the class variable and individual years of yield data as well as mean relative yield were used as the response variable. Yield variance reduction within a given field resulting from partitioning observations by soil type was calculated following Eq. [1] to [3].

The association between MZ Analyst-derived MZs based on crop yields and soil map units from the published, order two county soil survey (scale 1:15 840) (Neely, 1987) was assessed using a mean score statistics approach with modified ridit scores in a stratified Mantel-Haenszel analysis (FREQ, SAS Inst., 2000). For each field, the row mean scores statistic (Q_{SMH}) was used to test for (i) differences among soil types in yield level MZs while controlling for crop species, and (ii) differences between crop species in yield level MZs while adjusting for the effects of soil type. Cramer's V was used to measure general association among variables within each stratum of the analysis. The Jonckheere-Terpstra statistic was used to detect ordered differences among yield levels within each stratum.

Finally, the success with which a soil type could be identified from yield data was characterized with discriminant analysis. For each field, soil type was the class variable and the relative yield data from all study years were the quantitative variables. Stepwise discriminant analysis identified the years of yield data that were significant indicators of differences among soil types (STEPDISC, SAS Inst., 2000). These significant quantification variables were then used in discriminant analysis to estimate probable error for identifying soil map units from yield data (DISCRIM, SAS Inst., 2000). To avoid bias in error count estimates, the dataset for a given field was divided into two subsets. All observations associated with a specific, georeferenced location were randomly assigned to either the classification dataset or the validation dataset. The classification dataset was used to derive the discriminant function and test it with cross validation while the validation dataset was used for a subsequent, independent test of the discriminant function. The prior probability of an observation belonging to a given soil map unit was proportional to the map unit area within the field.

RESULTS AND DISCUSSION

Experimental Site

The county soil survey identified three major soil series in the study fields (Neely, 1987). Blount silt loams (fine, illitic, mesic Aeric Ochraqualfs) are somewhat poorly drained soils found on broad, nearly level areas (0–1% slopes). Glynwood silt loams (fine, illitic, mesic Aquic Hapludalfs) are moderately well drained, eroded soils found on knolls and short sideslopes with slopes of 1 to 4%. Pewamo silty clay loams (fine, mixed, mesic Typic Argiaquolls) are very poorly drained, nearly level soils found in depressions and natural drainageways. Field D was dominated by Blount soils while the remaining fields were dominated by Pewamo soils. Percentage land areas mapped to Blount soils were 59, 39, 41, and 31% while percentage land areas mapped to Pewamo were 41, 59, 48, and 60% for D, F, R, and V, respectively. Glynwood soils were minor in all study fields, mapping to <0.5, 1, 10, and 9% of the land area in D, F, R, and V, respectively.

According to the soil survey, expected corn productivities under a high level of management are 6.0, 6.7, and 7.8 Mg ha^{-1} for Glynwood, Blount, and Pewamo soils, respectively (Neely, 1987). Expected soybean productivities are 2.0, 2.4, and 2.8 Mg ha^{-1} for Glynwood,

Table 1. Selected univariate statistics for corn and soybean yield monitor data from experimental fields.

	Field D, yield†					Field F, yield					Field R, yield					Field V, yield				
Year	Crop‡	Avg.	Max	IQR	CV	Crop	Avg.	Max	IQR	CV	Crop	Avg.	Max.	IQR	CV	Crop	Avg.	Max.	IQR	CV
		—— Mg ha^{-1} ——			%		—— Mg ha^{-1} ——			%		—— Mg ha^{-1} ——			%		—— Mg ha^{-1} ——			%
1996	C	5.67	8.47	2.29	22.1	C	3.74	5.47	1.53	22.3	S	2.18	3.15	0.99	19.4	S	2.45	3.29	0.94	18.6
1997	W	–§	–	–	–	C	4.57	8.45	2.82	27.7	C	7.63	11.01	3.33	19.8	C	7.36	10.21	2.71	17.3
1998	W	–	–	–	–	S	2.76	3.94	1.22	12.8	C	7.81	11.06	1.59	14.8	C	8.17	11.49	1.31	12.4
1999	C	8.97	11.54	2.03	6.8	C	10.43	13.08	2.73	6.8	S	2.28	3.37	1.11	24.9	S	2.45	3.28	0.86	19.4
2000	S	3.39	4.91	1.41	13.0	S	3.31	5.36	1.29	17.6	C	7.37	11.50	4.14	26.2	C	8.48	11.33	2.48	13.6
2001	C	–	–	–	–	C	–	–	–	–	S	3.39	4.16	0.75	7.2	S	2.98	4.02	0.97	14.9
2002	S	2.67	3.87	1.08	18.3	S	3.08	3.98	0.90	15.4	C	–	–	–	–	C	–	–	–	–
All corn		7.32	10.00	2.16			6.25	9.00	2.36			7.60	11.19	3.02			8.00	11.01	2.17	
All soybean		3.03	4.39	1.25			3.05	4.43	1.14			2.62	3.56	0.95			2.63	3.53	0.93	

† Within a field-year, average (Avg.) yield is the mean of all observations following rectification of yield monitor data; maximum (Max) yield is the mean of all observations in the top fifth percentile, IQR is the difference between the yield values observed at the 75th and 25th quartile divisions, and CV is the coefficient of variation.
‡ C, S, and W represent corn, soybean, and wheat, respectively, the crop grown in a given year.
§ No data indicates no yield monitor data for corn or soybean available in that year.

Blount, and Pewamo soils, respectively. The soil survey capability classification identifies Glynwood soils as class IIIe, indicating that this soil has severe limitations to plant productivity associated with erosion. Blount and Pewamo soils are classified as IIw, indicating moderate limitations associated with wetness that can be at least partially corrected with artificial drainage. All fields in this study have subsurface tile drains but the location and intensity of the drains was not known.

Crop Productivity

Whole-field average corn yields ranged from a study low of 3.7 Mg ha^{-1} observed in 1996 in Field F to a high of 10.4 Mg ha^{-1} observed in 1999, also in Field F (Table 1). Within-field maximum corn yields were observed to range from 5.5 to 13.1 Mg ha^{-1} (F, 1996 and 1999, respectively). For soybean, whole-field average yields ranged from 2.2 Mg ha^{-1} (R, 1996) to 3.4 Mg ha^{-1} (D, 2000 and R, 2001) while whole-field maximum yields ranged from 3.2 Mg ha^{-1} (R, 1996) to 5.4 Mg ha^{-1} (F, 2000). On average, higher corn yields were observed in R and V (7.6 and 8.0 Mg ha^{-1}, respectively) when compared with D and F (7.3 and 6.3 Mg ha^{-1}, respectively). Conversely, R and V averaged 2.6 Mg soybean ha^{-1} while D and F averaged 3.0 Mg soybean ha^{-1}. Mean relative yields were fairly consistent among crop species and study fields, ranging from 67 to 74% of the maximums.

Correlation analyses within a given field indicate both some degree of temporal stability in spatial yield patterns as well as some species–specific differences in these patterns. Among the four study fields, there were 51 paired comparisons between individual years of yield data (six within-field comparisons in D with 4 yr of yield data, and 15 in F, R, and V with 6 yr of yield data each) resulting in 41 significant positive correlations (Pearson correlation, r, 0.06 to 0.75; $P < 0.05$), four significant negative correlations ($r = -0.12$ to -0.43; $P < 0.001$), and six nonsignificant relationships (data not shown). Five of the six nonsignificant relationships were observed in cross-species comparisons but the negative correlations were divided among inter- and intra-species comparisons (two cases each). Field V was the most consistent with only positive, significant correlations observed between individual years, irrespective of crop species ($r = 0.06$–0.69 and $r = 0.16$–0.68 for inter and intra-species correlations, respectively). Field F was the least consistent with seven significant positive, three significant negative, and five nonsignificant correlations observed.

When individual years of corn yield data were correlated with the multi-year average for corn in that field, r ranged from 0.56 to 0.95 ($P < 0.001$) (Table 2). Likewise, when individual years of soybean yield data were correlated with the multi-year average for soybean in that field, r ranged from 0.48 to 0.89 ($P < 0.001$). These results are similar to Lamb et al. (1997), who observed

Table 2. Pearson correlation coefficients for relationships between relative yields of individual years and the mean relative corn yield, soybean yield and combined corn and soybean yields.

	Field D, mean relative yields				Field F, mean relative yields				Field R, mean relative yields				Field V, mean relative yields			
Year	Crop†	Corn	Soy	Both	Crop	Corn	Soy	Both	Crop	Corn	Soy	Both	Crop	Corn	Soy	Both
1996	C	0.95***	0.23***	0.75***	C	0.60***	0.41***	0.66***	S	0.24***	0.67***	0.47***	S	0.38***	0.79***	0.66***
1997	W	–‡	–	–	C	0.69***	−0.23***	0.30***	C	0.63***	0.63***	0.69***	C	0.74***	0.68***	0.76***
1998	W	–	–	–	S	0.04NS	0.68***	0.42***	C	0.76***	0.40***	0.65***	C	0.78***	0.51***	0.68***
1999	C	0.56***	0.33***	0.55***	C	0.71***	0.42***	0.73***	S	0.70***	0.81***	0.82***	S	0.69***	0.89***	0.87***
2000	S	0.55***	0.61***	0.72***	S	−0.18***	0.48***	0.15***	C	0.85***	0.47***	0.75***	C	0.79***	0.41***	0.62***
2001	C	–	–	–	C	–	–	NS	S	0.37***	0.50***	0.47***	S	0.70***	0.80***	0.82***
2002	S	−0.16NS	0.83***	0.49***	S	0.50***	0.61***	0.71***	C	–	–	–	C	–	–	–
							Mean relative yields									
Soy only		0.30***				0.21***				0.67***				0.71***		
Both		0.82***	0.79***			0.82***	0.73***			0.93***	0.90***			0.90***	0.94***	

*** Correlations are significant ($p < 0.001$). NS, not significant.
† C, S, and W represent crop species corn, soybean, and wheat, respectively.
‡ No data indicates no yield monitor data for corn or soybean available in that year.

better correlations between individual year corn yield and the multi-year corn yield average than among individual years of corn yield. In our study, when cross-species correlations were performed, *r* values decreased when compared to intra-species correlations. The *r* values ranged from −0.18 to 0.70 and −0.23 to 0.68 for relationships between individual years of soybean data and multi-year corn averages and between individual years of corn data and multi-year soybean averages, respectively. These correlation results suggest that there may be some differences in yield limiting factors for the two crop species in the rotation.

Zone Delineation and Performance Measures

Before a clustering algorithm can be applied to data, a measure of similarity must be selected that provides the foundation for a decision rule regarding the assigning of a particular observation to the domain of a given cluster. When multiple independent variables are used to identify clusters, attributes of the input data layers that require consideration include correlation and equality of variance (Fridgen et al., 2004). While Pearson correlations indicated that within a given field some or all of the yield data from individual years were significantly correlated, Levine's test for homogeneity of variance found unequal variances for at least one pair of yield comparisons per field. Therefore, the Malalanobis distance measure of similarity was chosen. This measure can account for both correlation between independent variables and unequal independent variable variances (McBratney and Moore, 1985; Odeh et al., 1992).

For both the FPI and the NCE relative minimum values indicate optimal clustering. As described in detail by Bezdek (1981), Odeh at el. (1992), and Boydell and McBratney (2002) and summarized by Fridgen et al. (2004), a minimum FPI represents the least amount of membership sharing of observations among clusters as a result of the clustering order, while the relative minimum in the NCE represents the attainment of the greatest amount of organization. In this experiment, the two indices did not appear to provide complimentary information in the identification of the optimum number of MZs for any of the study fields (Fig. 1 and Fig. 2). For CYB MZs in D, F, and V, the FPI indicated optimal clustering with four MZs (Fig. 1a). In R, the FPI for CYB MZs reached a local minimum at four MZs but the global minimum occurred at six MZs, the maximum number of MZs evaluated. In contrast, the NCE values indicated that two MZs optimized the data clustering for CYB MZs, especially in D and R (Fig. 2a). In F and V, NCE values with four CYB MZs were only slightly greater than with two MZs indicating that two and four MZs resulted in comparable data organization. Results were similar for SYB and YB MZs. For SYB MZs, FPI indicated optimal data clustering occurred with four (R and V), five (D), or six (F) zones but NCE global minima were observed at two (F, R, and V) or three (D) MZs. Global FPI minima for YB MZs were achieved with three (D), five (V), or six (F and R) MZs but NCE global minima occur at two MZs in all fields although local NCE minima were observed at five (D and V) and six (F) MZs.

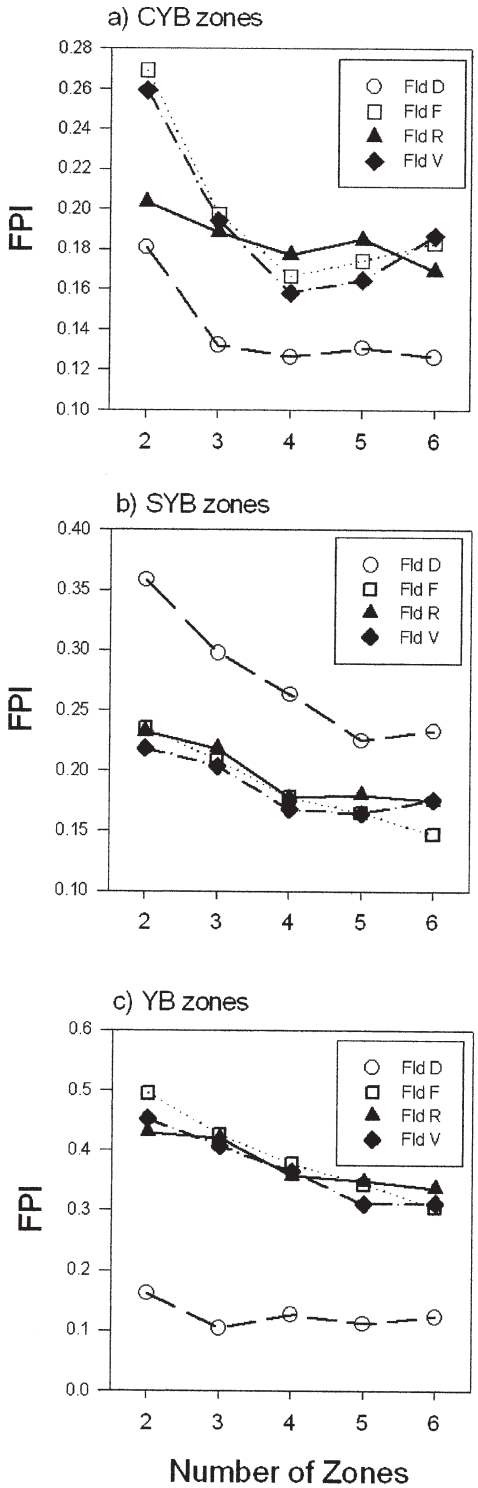

Fig. 1. Fuzziness performance index (FPI) values shown as a function of the number of management zones (MZs) for (a) corn yield-based (CYB) MZs, (b) soybean yield-based (SYB) MZs, and (c) combined crop yield-based (YB) MZs.

Identifying the most appropriate number of clusters is a noted difficulty in the interpretation of unsupervised clustering results (Afifi and Clark, 1984). In the MZ Analyst case study (Fridgen et al., 2004), the FPI and

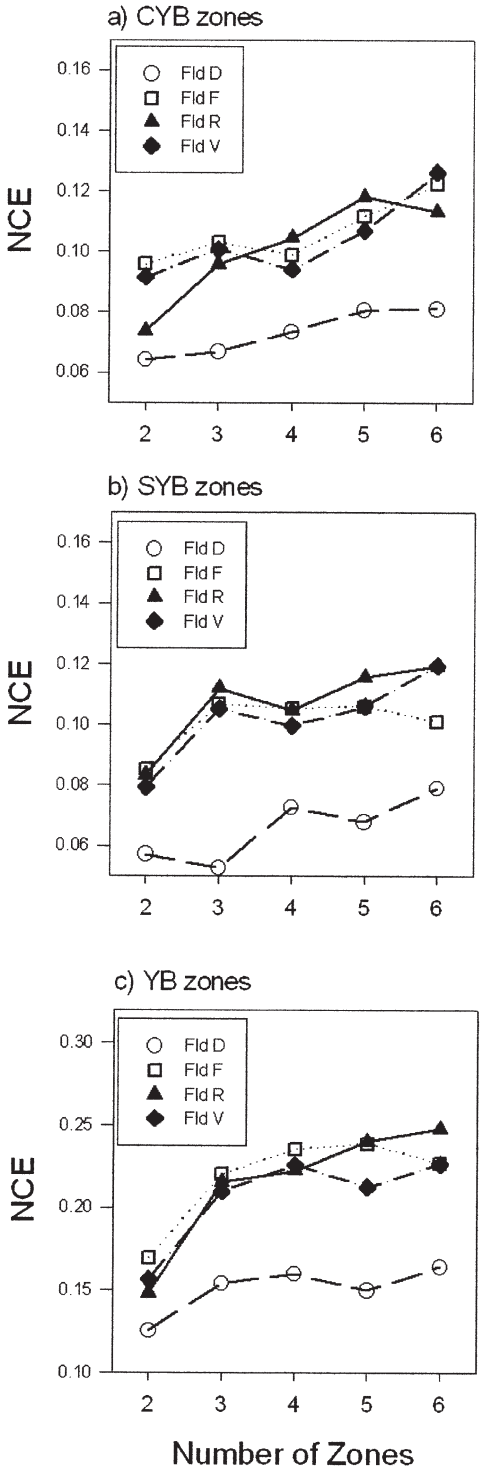

Fig. 2. Normalized classification entropy (NCE) values shown as a function of the number of management zones (MZs) for (a) corn yield-based (CYB) MZs, (b) soybean yield-based (SYB) MZs, and (c) combined crop yield-based (YB) MZs.

NCE behaved similarly as a function of the number of clusters examined, converging on an optimum clustering order solution. For FPI and NCE, the theoretical expectation is that both indices will decrease at the same rate until the optimum combination of characteristics is described by the clusters (Boydell and McBratney, 2002). However, in their study on spatial structure in cotton (*Gossypium hirsutum* L.) yields, these authors observed that FPI and NCE suggested more than one order of clustering may be optimal and the uncertainty was between cluster orders that were very different (2 vs. 7 or 9 clusters). The authors noted this result further contradicted the literature, which indicated that uncertainty, if it exists, should be between sequential clusters (e.g., 2 vs. 3) and not between isolated clusters.

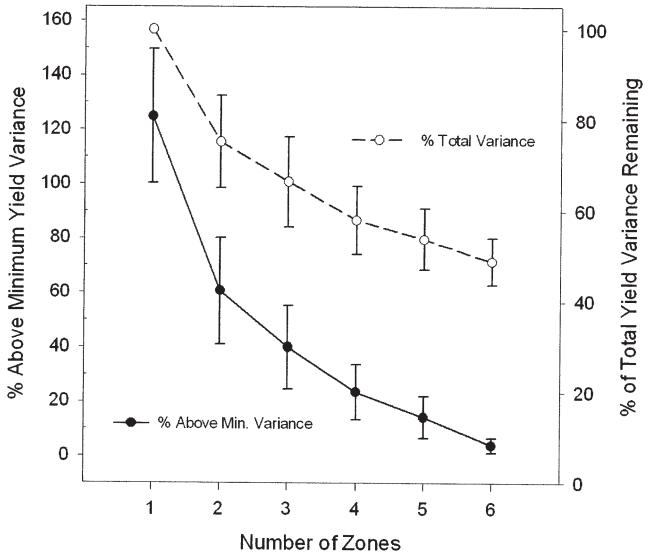

Fig. 3. Two measures of yield variance reduction resulting from delineation of experimental fields into two to six zones using both corn and soybean yield data. Data points are the means for both crops in all experimental fields. Error bars show the 95th percent confidence interval.

Regardless, neither the Boydell and McBratney (2002) nor the Fridgen et al. (2004) studies provide guidance on the degree of change of an index value that should be considered an indication of meaningful difference in cluster performance. Fridgen et al. (2004) report changes in FPI and NCE values of ≤0.1 and ≤0.03, respectively, while Boydell and McBratney (2002) observed changes in FPI and NCE of ≤0.03 and ≤0.04, respectively. We observed changes in FPI ranging from 0.05 to 0.13 (Fig. 1) and changes in NCE ranging from 0.02 to 0.1 (Fig. 2). Presumably, both the degree of change in index value and the absolute value are relevant to the interpretation of clustering success. These topics have yet to be addressed in published reports on clustering algorithm applications to agricultural data making these indicies difficult to interpret for practical purposes.

The degree to which a number (n) of clusters reduces the within-cluster variability of spatial attributes when compared with $n - 1$ clusters has been proposed as an additional method by which to evaluate classification success (Fridgen, 2000; Fraisse et al., 2001). For the three MZ delineation strategies we examined, the percentage above minimum variance and the percentage of total variance tended to decline with increasing clustering order. In all cases, general minima were attained with six MZs, the highest order of clustering examined, but most pronounced changes occurred with delineations of two, three, and four MZs (Fig. 3). In the case of YB MZs,

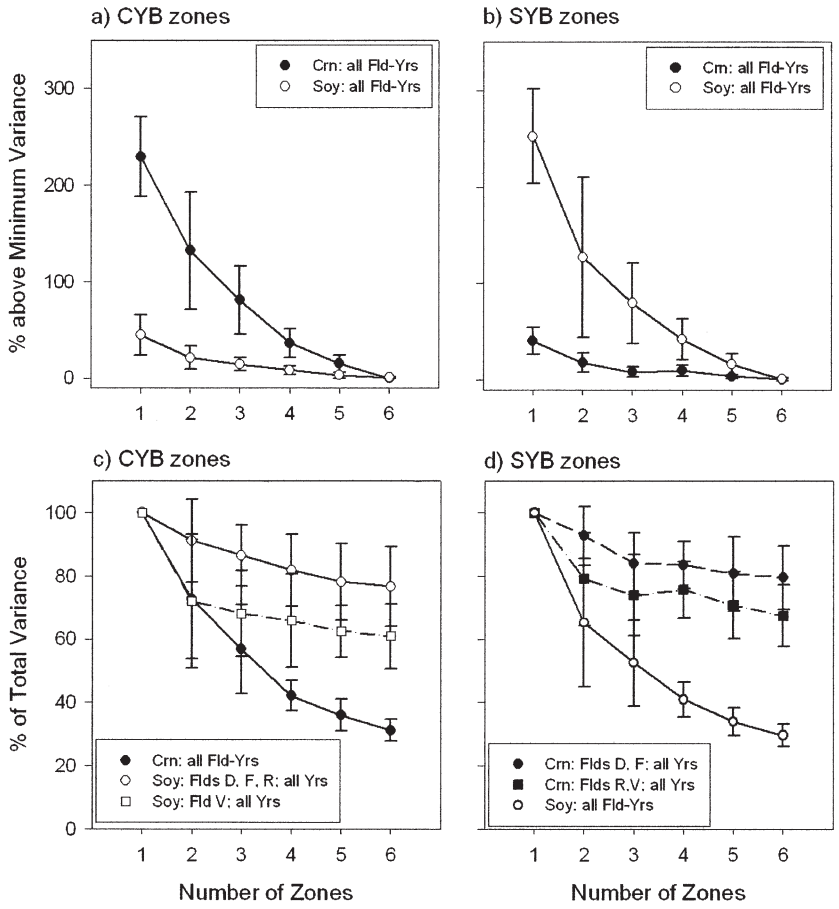

Fig. 4. The two measures of yield variance reduction for (a, b) corn yield-based (CYB) management zones (MZs), and (c, d) soybean yield-based (SYB) MZs shown as a function of the number of MZs. Data points are crop-specific means for the experimental fields that did not differ significantly. Error bars show the 95th percent confidence interval.

ANOVA found no significant effect of crop species or field on either variance reduction parameter but classification orders up to four MZs significantly effected variance reduction. Whole-field variances (one MZ) averaged 125% above the minimum variance while four, five, and six MZ variances averaged 23, 13, and 4% above the minimum variance, respectively. Four MZs reduced variation to 58% of the original total variance, which was not significantly different from the reduction in total variance resulting from delineation into six MZs.

Examination of variance reduction as a function of clustering order within CYB and SYB MZs also found four MZs to account for the majority of the variance reduction but results showed some degree of crop specificity in the MZs delineated. In CYB MZs, whole-field variances averaged 230 and 45% above minimum variances for corn and soybean, respectively (Fig. 4a). Conversely, in SYB MZs, whole-field variances averaged 35 and 250% above minimum variances for corn and soybean, respectively (Fig. 4b). These results reflect the significantly higher minimum variances for the alternate crop when compared with the delineating crop in a crop-specific MZ delineation strategy. Likewise, for both corn within CYB MZs and soybean within SYB MZs, delineation into six MZs reduced the variance to approximately 30% of the original whole-field variances (Fig. 4c and 4d). However, a six-MZ delineation reduced alternate crop variances to only 60 to 80% of the original whole-field variances.

As with FPI and NCE, the literature offers little guidance on the interpretation of variance reduction values for determining the relative success of different clustering orders. Fridgen (2000) did not consider MZs to account for the variability in a classification attribute unless the S_T^2 was reduced at least 10% when compared with the whole-field variance. By this criterion, three or more MZs were sufficient in our study, irrespective of classification strategy (Fig. 4c and 4d). Fridgen (2000) also proposed that within-MZ variance reductions to within 10% of the minimum as an indication of optimum clustering. However, both Fridgen (2000) and Fraisse et al. (2001) observed that variance reduction as a function of clustering order varied among years, strongly influenced by weather and crop growth conditions. Our approach of using ANOVA to identify the lowest number of MZs where response variables were not significantly different from results with maximum number of MZs evaluated permits this year-to-year variability to be taken into account.

From a practical perspective, an added consideration for the selection of the optimum clustering is the land area of the smallest MZ. With ≤4 clusters, the minimum

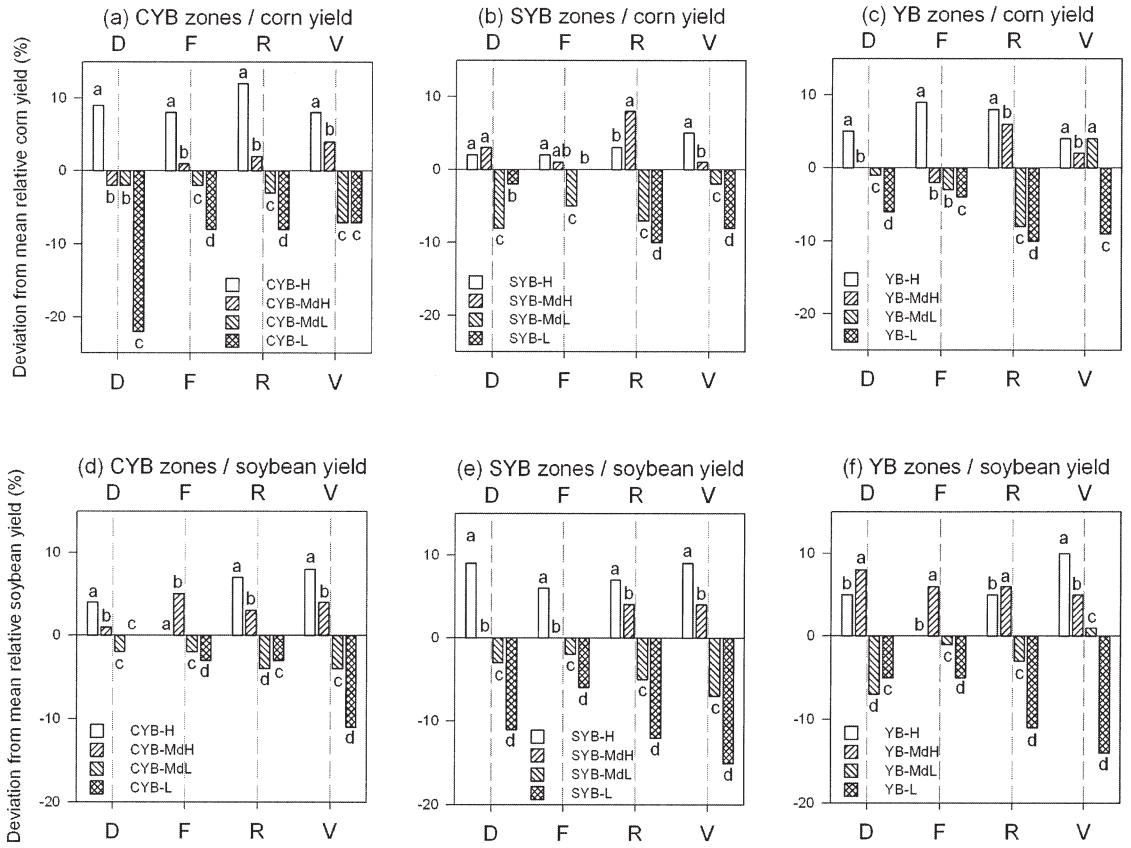

Fig. 5. Deviation of management zone (MZ) mean relative yield from whole field means following delineation of fields into four MZs. Data are show for (a–c) corn and (d–f) soybean for the three zone delineation strategies including (a, d) corn yield-based (CYB) MZs, (b, e) soybean yield-based (SYB) MZs, and (c, f) combined crops yield-based (YB) MZs. Within a field, columns topped with the same letter are not significantly different ($p > 0.05$).

size of the smallest MZ within any field generally exceeded 15% of the land area (data not shown). The only exception to this observation occurred in F where identifying four CYB MZs resulted in a MZ that comprised only 5.5% (0.62 ha) of the field area. However, when fields were delineated into five or six MZ, the smallest MZs were <16% of the field land area. For six CYB MZs, the smallest MZs were 0.23, 0.56, 0.17, and 1.95 ha in D, F, R, and V, respectively. Likewise, for six SYB MZs, the smallest MZs were 0.10, 0.63, 1.45, and 1.37 ha in D, F, R, and V, respectively.

For precision agriculture management objectives that are linked to differences in crop productivity potential, it is presumably important to assess the yield differences among the different clusters at the optimal clustering order. For corn within four CYB MZs and soybean within four SYB MZs, mean relative yields in highest and lowest yielding MZs were significantly different from each other (Fig. 5a and 5e). Highest and lowest yielding MZs were at least 6% above and below field relative means, respectively. With one exception, these MZs were also significantly different in mean relative yield from the intermediate MZs indicating that, on average, the four MZs within a field were distinct from each other. It should be noted, however, that the yield ranking of MZs was not entirely stable among the individual years for either crop. For example, exact matches between CYB MZ yield rank in an individual year with the CYB MZ yield rank based on mean relative corn yield occurred in 58% of the comparisons (data not shown). The remaining comparisons generally involved mismatches of just one rank level but every field except V had one comparison with a mismatch of two rank levels. Results were similar for soybean within SYB MZs.

Association and Agreement among Delineation Strategies

The ANOVA of soybean yields within CYB MZs, corn yields within SYB MZs, and yields of both crops within YB MZs again identified crop specificity in zone delineation. Mean relative yields of the alternate crop within CYB and SYB MZs were less differentiated from each other when compared with the yield differences of the MZ-delineating crop. Furthermore, alternate crop MZ rankings did not always match the MZ rankings of the MZ-delineating crop. For example, in the lowest yielding SYB MZ in D, the mean relative soybean yield was 11% below the field mean soybean yield (Fig. 5e) but the mean relative corn yield was only 2.5% below the field mean corn yield (Fig. 5b). For the corn yields this MZ ranked third and not fourth.

Overall, MZs appeared relatively more crop-specific in D and F when compared to R and V where alternate crops appeared to have greater yield differentiation within a crop-specific MZ (Fig. 5). This observation is supported

Table 3. Degree of association among management zone (MZ) delineation strategies including corn yield-based (CYB) MZs, soybean yield-based (SYB) MZs, and combined corn and soybean yield-based (YB) MZ. Statistics characterize general association (Cramer's V) and linear (Pearson) correlation of the zone ranks.

Association statistic	No. of MZs	CYB vs. SYB				YB vs. CYB				YB vs. SYB			
		Field D	Field F	Field R	Field V	Field D	Field F	Field R	Field V	Field D	Field F	Field R	Field V
Cramer's V	2	0.20	0.03†	0.61	0.62	0.62	0.02†	0.92	0.75	0.09	0.72	0.68	0.87
	3	0.31	0.30	0.40	0.47	0.47	0.73	0.50	0.83	0.55	0.42	0.53	0.56
	4	0.29	0.32	0.37	0.38	0.38	0.73	0.50	0.60	0.63	0.42	0.68	0.51
	5	0.36	0.29	0.38	0.39	0.48	0.68	0.55	0.55	0.49	0.44	0.62	0.63
	6	0.39	0.28	0.38	0.38	0.48	0.62	0.54	0.56	0.55	0.46	0.57	0.57
Pearson correlation	2	0.20	0.03‡	0.61	0.62	0.62	0.02‡	0.92	0.75	0.09	0.72	0.68	0.87
	3	0.07	0.11	0.56	0.48	0.12	−0.18	0.63	0.86	0.73	0.51	0.66	0.54
	4	0.10	0.26	0.40	0.48	0.16	0.62	0.61	0.53	0.63	0.41	0.62	0.69
	5	0.34	0.20	0.50	0.56	0.39	0.56	0.62	0.62	0.62	0.36	0.68	0.77
	6	0.29	0.17	0.53	0.52	0.47	0.64	0.62	0.61	0.69	0.35	0.62	0.69

† Reject alternative hypothesis of general association.
‡ Reject alternative hypothesis of non-zero correlation.

by the analysis of degree of association among MZ delineation strategies at each clustering order. With the exception of the two-MZ case in F, chi square analysis found significant general association between CYB and SYB MZ and between either CYB or SYB MZs and YB MZs (Table 3). For clustering orders ≥4, the Cramer's V statistic resulting from CYB-SYB comparisons were comparable among the fields, ranging between 0.29 and 0.39. These values for Cramer's V, which can vary from 0 (no association) to 1 (perfect association), indicate a moderate CYB-SYB level of association in all fields. However, *r* values characterizing the linearity of the relationship indicate that when the relative ranking of the MZs is considered SYB-CYB associations are stronger in R and V when compared to D and F. For the four-MZ delineation, *r* values were 0.40 and 0.48 in R and V, respectively, vs. 0.10 and 0.26 in D and F, respectively.

Agreement, a special case of association, was also stronger between CYB and SYB MZs in R and V than in D and F. While P_0 values were not necessarily different among the fields for clustering orders ≥4, K_w values indicate stronger agreement among delineation strategies for R and V (Table 4). The K_w can range from −1 to +1 with negative and positive values indicating agreement of less than and greater than chance, respectively. Stokes et al. (2000) characterize $0 < K_w \leq 0.4$, $0.4 \leq K_w < 0.8$, and $0.8 \leq K_w \leq 1.0$ as indicative of slight, moderate, and excellent agreement, respectively. According to this rating, agreement observed between CYB and SYB MZs in R and V was moderate at all clustering orders while agreement in D and F was only slight. Likewise, agreement between CYB or SYB MZs and YB MZs was moderate to excellent in R and V but only slight to moderate in D and F.

Lower *r* and K_w values but similar Cramer's V values for strategy comparisons in D and F when compared with R and V suggest that association among strategies in D and F occurred in cells further from contingency table diagonals. For the four-MZ comparison of CYB to SYB MZs, examination of cell chi-square values (data not shown) and observation distributions in the contingency table illustrates this effect (Fig. 6). In Field V, >75% of the observations in the highest yielding CYB MZ corresponded to the highest or medium-highest yielding SYB MZs (Fig. 6d). Likewise, 80% of the observations in the lowest yielding CYB MZ corresponded to lowest or medium-lowest yielding SYB MZs. Results were similar for CYB MZ rank distributions within SYB MZs. In contrast, in D the lowest yielding MZ was the dominant SYB MZ within the second highest yielding CYB MZ; the second highest MZ was the dominant CYB MZ within the lowest yielding SYB MZ (Fig. 6a).

For most clustering orders in all fields, both association and agreement between CYB or SYB MZs and YB MZs were stronger than the association and agreement observed between CYB and SYB MZs. This result is expected as both crop species were contributing to the delineation of the MZs. Indeed, stepwise discriminant analysis found that for two, four, and six YB MZs both crops and almost all years of yield data contributed significantly to the delineation of the zones (results not shown; three and five YB MZs not evaluated). The sole

Table 4. Agreement among management zone (MZ) delineation strategies including corn yield-based (CYB) MZs, soybean yield-based (SYB) MZs, and combined corn and soybean yield-based (YB) MZs. Statistics are for raw (P_O) and weighted (K_W) agreement.

Agreement statistic	No. of MZs	CYB vs. SYB				YB vs. CYB				YB vs. SYB			
		Field D	Field F	Field R	Field V	Field D	Field F	Field R	Field V	Field D	Field F	Field R	Field V
P_O	2	0.60	0.50	0.81	0.82	0.81	0.51	0.96	0.88	0.55	0.85	0.84	0.94
	3	0.34	0.33	0.52	0.62	0.17	0.14	0.51	0.89	0.63	0.56	0.65	0.67
	4	0.20	0.26	0.32	0.40	0.40	0.79	0.30	0.62	0.26	0.31	0.53	0.54
	5	0.42	0.30	0.31	0.38	0.43	0.43	0.43	0.53	0.36	0.26	0.28	0.62
	6	0.22	0.25	0.19	0.35	0.33	0.57	0.27	0.26	0.32	0.31	0.41	0.46
K_W	2	0.20	0.02†	0.61	0.62	0.62	0.02†	0.92	0.75	0.09	0.70	0.67	0.87
	3	0.04	0.05	0.45	0.42	−0.07	−0.24	0.45	0.85	0.59	0.42	0.57	0.52
	4	0.06	0.16	0.27	0.34	0.15	0.66	0.38	0.50	0.36	0.27	0.49	0.53
	5	0.28	0.17	0.34	0.41	0.30	0.44	0.47	0.51	0.36	0.22	0.44	0.64
	6	0.02†	0.14	0.31	0.37	0.31	0.56	0.42	0.41	0.45	0.26	0.44	0.52

† K_W not significantly different from 0.

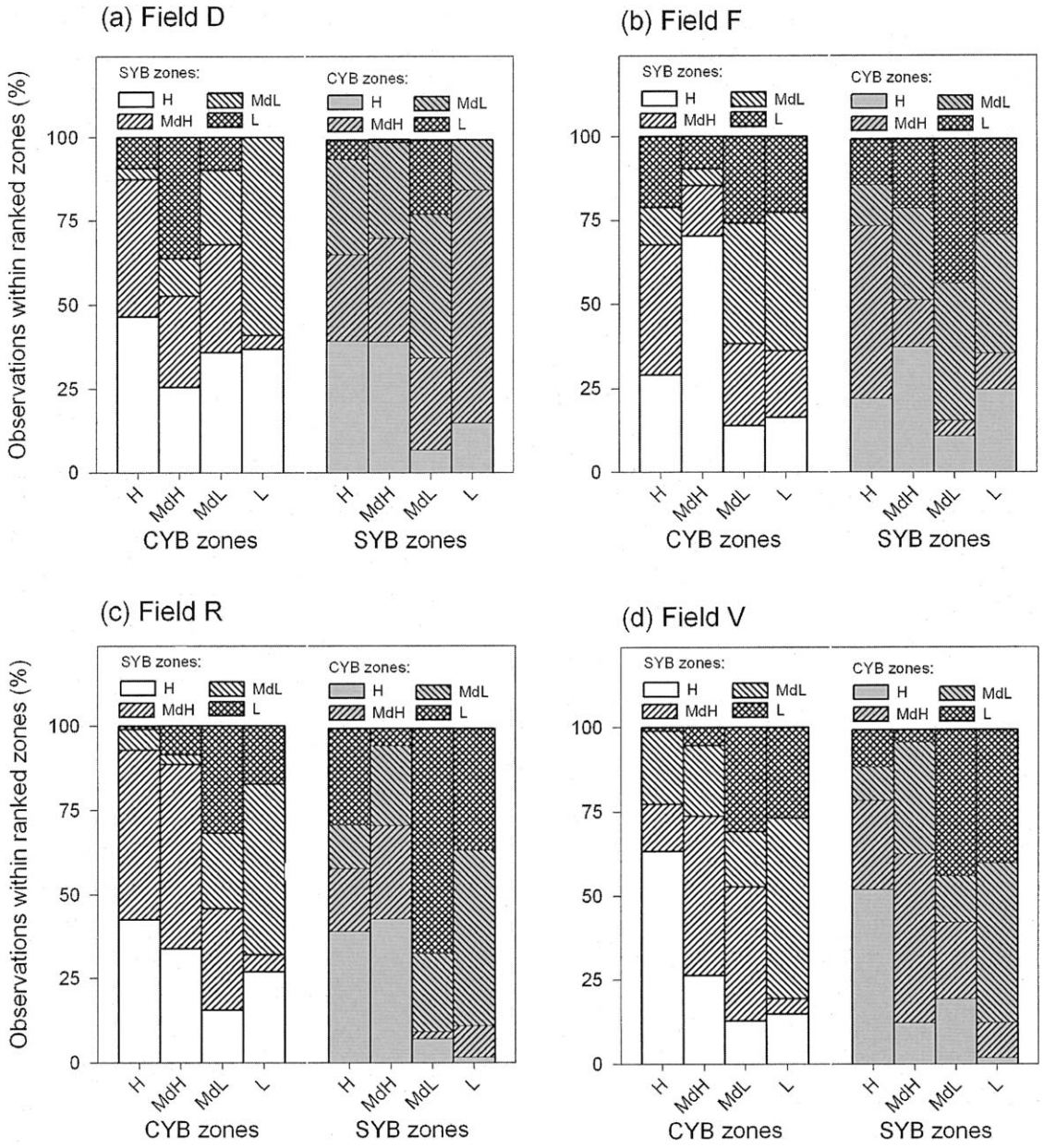

Fig. 6. Distribution of soybean yield-based (SYB) management zone (MZ) rankings within each rank of corn yield-based (CYB) MZs and distribution of CYB MZ rankings within each rank of SYB MZs. Data are shown for four MZ delineations in each field; stacked bars show percent of observations within a MZ ranking that were identified as high (H), medium high (MdH), medium low (MdL), and low (L) yield by the alternate delineation strategy.

exception occurred in delineating two YB MZs in F where 1999 corn did not contribute significantly to the MZ discrimination process. It is interesting to note that this year of corn data represented the highest whole-field average yield (10.43 Mg ha^{-1}) observed both in F and among the other experimental fields as well (Table 1).

To date, the literature contains no known examples of comparisons of corn and soybean productivity MZs developed using unsupervised clustering. Our results suggest that spatial differences exist in the yield limiting factors for these two crops. Thus, zone management of factors that are specific to only one crop in the rotation such as N may require different MZs than factors that are managed for both crops in the rotation (e.g., P or K). We note, however, that our MZs were developed using a very limited database. Lamb et al. (1997) examined year-to-year correlations in corn yields and suggested that more than 5 yr may be needed to identify stable yield patterns. Recent research on cotton found that a minimum of 5 yr of yield monitor data was required to identify stable MZs (Boydell and McBratney, 2002). In this study, we had insufficient data to evaluate the number of years of yield monitor data required to identify temporal stability for CYB, SYB, and YB MZs.

Yield-Based Zones and Soil Series

The ANOVA of yields using soil series as the class factor revealed significant differences in three of the

four fields. Mean relative corn and soybean yields were highest in the Pewamo silty clay loam soils (Table 5). On Pewmo soils, corn yields averaged 72.7, 74.3, and 78.3% of maximum or 6.5, 8.3, and 8.6 Mg ha^{-1} in F, R, and V, respectively. Pewamo soybean yields were 74.7, 78.4, and 80.9% of maximum or 3.3, 2.8, and 2.9 Mg ha^{-1} in F, R, and V, respectively. In Fields F and R, Glynwood soils were lowest yielding (mean of 5.7 and 7.0 Mg corn ha^{-1} and mean 3.1 and 2.4 Mg soybean ha^{-1}, respectively) with Blount soils yielding at an intermediate level (mean of 6.3 and 7.5 Mg corn ha^{-1} and mean 3.1 and 2.5 Mg soybean ha^{-1}, respectively). In V, Blount soils averaged lowest yields (7.6 Mg corn ha^{-1} and 2.3 Mg soybean ha^{-1}) while Glynwood soils averaged intermediate yields (7.9 Mg corn ha^{-1} and 2.5 Mg soybean ha^{-1}). In D, soil-specific corn yields were not significantly different; soybean yields were greater on the Blount compared with Pewamo soil but the magnitude of the difference in mean relative yield was small (1%).

It is important to recall that Glynwood soils were minor in all fields. In D, the Glynwood land area was so small that we omitted it from all analyses. Thus, the advantage from partitioning by soil series in these fields must be associated with differences between Blount and Pewamo. In R and V, the differences in yield between Blount and Pewamo soils are comparable to the magnitude predicted by the soil survey (1.1 Mg corn ha^{-1} and 0.4 Mg soybean ha^{-1}; Neely 1987), but soil-specific yield differences were less than expected in F and inconsequential in D. Regardless, in all fields, the soil-specific deviations from mean relative yields tended to be smaller than the deviations observed for highest and lowest yielding MZ Analyst-derived MZs, especially when considering corn in CYB MZs and soybean in SYB MZs (four clusters; Fig. 5). For example, in V the soybean yields were 10.0% below and 6.2% above the mean relative yield for Blount and Pewamo, respectively. For MZ Analyst-derived MZs (four MZs), lowest yielding MZs for soybean were 11 to 15% below the mean relative yield in CYB and SYB strategies, respectively. Highest yielding MZs for soybean were 8 to 10% above the mean relative yields for CYB and SYB or YB strategies, respectively.

Likewise, the variance reduction following partitioning mean relative yield observations by soil series indicates that soil series accounts for only a portion of the yield variability that was accounted for by yield-based MZs. The variance remaining following partitioning mean relative yield by soil series ranged from 74.3 to 100.0% for corn and 52.6 to 99.7% for soybean (Table 5). These variance reductions were less than the variance reductions observed following optimal MZ Analyst-delineated CYB and SYB MZs for a given field (Fig. 4). However, in R and V, yield variance reductions following partitioning by soil series did exceed the 10% criteria proposed by Fridgen (2000) for identifying significant clustering factors.

Analysis of association between CYB or SYB MZs and soil map units identified both soil series and crop species influences. By controlling for crop species effects in the analysis of the distribution of MZ yield ranks

Table 5. Mean relative yields and coefficients of variation for soil series within fields. The percent variance remaining is shown for both the mean relative yield and the mean of the individual years.

Soil type	Statistic	D			F			R			V		
		Corn	Soy	Both	Corn	Soy	Both	Corn	Soy	Both	Corn	Soy	Both
		%											
Blount	mean (CV)	75.5 (9.2)	71.0b (11.4)†	73.3 (8.2)	70.4b (9.7)	70.5b (9.1)	70.4b (7.3)	66.7b (16.1)	71.0b (12.1)	68.9b (13.2)	69.4c (10.8)	64.7c (13.9)	67.0c (11.2)
Glynwood	mean (CV)	–	–	–	63.3c (4.3)	69.4b (7.5)	66.4c (5.0)	62.5a (13.2)	66.3 (12.4)	64.4c (12.2)	71.9b (11.0)	71.5b (15.2)	71.7b (12.5)
Pewamo	mean (CV)	75.1 (14.4)	72.0a (11.6)	73.5 (10.7)	72.7a (11.5)	74.7a (7.9)	73.7a (7.4)	74.3a (12.4)	78.4 (8.3)	76.4a (8.6)	78.3a (8.7)	80.9a (8.3)	79.6a (7.3)
All soils	mean (CV)	75.3 (11.6)	71.4 (11.5)	73.4 (9.3)	71.6 (10.9)	72.8 (8.7)	72.3 (8.9)	70.2 (15.2)	74.4 (11.7)	72.3 (8.9)	74.8 (11.1)	74.7 (14.6)	74.8 (8.4)
						% Variance remaining‡							
Mean relative yield		100.0	99.7	100.1	97.0	89.6	90.9	83.2	75.0	75.8	74.3	52.6	56.6
Mean individual years		99.5	96.2	97.9	94.4	90.8	92.6	89.0	83.0	86.0	82.3	67.3	74.7

† Relative yield values within columns followed by the same letter are not significantly different ($P > 0.05$).
‡ Calculated following Eq. [1].

Table 6. Association among rank of corn yield-based (CYB) management zones (MZs), rank of soybean yield-based (SYB) MZs, and soil series as characterized by Cramer's V, Jonckheere-Terpstra Z value (JT Z), and row mean scores (Q_{SMH}) statistics. Distribution of zone ranks within soil map units are shown in Fig. 7.

			Association			
Stratification	**Factor controlled**	**Statistic**	**Field D**	**Field F**	**Field R**	**Field V**
Crop species	**CYB**	**Cramer's V**	**0.23*****	**0.38*****	**0.36*****	**0.38*****
		JT Z	**−0.22NS**	**−11.07*****	**−27.06*****	**−34.58*****
	SYB	**Cramer's V**	**0.25*****	**0.38*****	**0.40*****	**0.44*****
		JT Z	**−2.90****	**−13.75*****	**−27.44*****	**−36.86*****
		Q_{SMH}	**1.64NS**	**308.30*****	**1665.29*****	**2584.43*****
Soil series	**Blount**	**Cramer's V**	**0.37*****	**0.09****	**0.44*****	**0.14*****
		JT Z	**0.81NS**	**−2.22***	**−16.56*****	**−2.89****
	Glynwood	**Cramer's V**	**NR†**	**NR**	**0.07NS**	**0.63*****
		JT Z	**NR**	**NR**	**0.31NS**	**−7.57*****
	Pewamo	**Cramer's V**	**0.11*****	**0.13*****	**0.21*****	**0.09*****
		JT Z	**−1.14NS**	**−5.65*****	**−9.65*****	**−5.67*****
		Q_{SMH}	**4.68***	**31.36*****	**265.48*****	**67.62*****

*** Statistical significance at $P < 0.05$. NS, not significant.**
**** Statistical significance at $P < 0.01$.**
***** Statistical significance at $P < 0.001$.**
† No result: insufficient observations to perform association statistics.

within a soil series, significant differences among soils were observed in F, R, and V (Q_{SMH} = 308.3, 1665.29, and 2584.43, respectively, $P < 0.001$; Table 6). In these fields, significant, negative Jonckheere-Terpstra Z values (Table 6) indicate Blount soils had more medium-low and low yielding CYB and SYB MZs when compared to Pewamo soils where high and medium-high yielding CYB and SYB MZs dominated (Fig. 7b–d). In D, distributions of CYB and SYB MZ ranks within Blount and Pewamo soils were similar (Fig. 7a). Controlling for soil series in the CYB-SYB comparative analysis of MZ rank distributions identified small but significant differences in all fields (Q_{SMH} 4.68–67.62, $P <$ 0.05). Significant Jonckheere-Terpstra Z values indicated these differences were ordered in F, R, and V. On Blount there were more than expected low and medium-

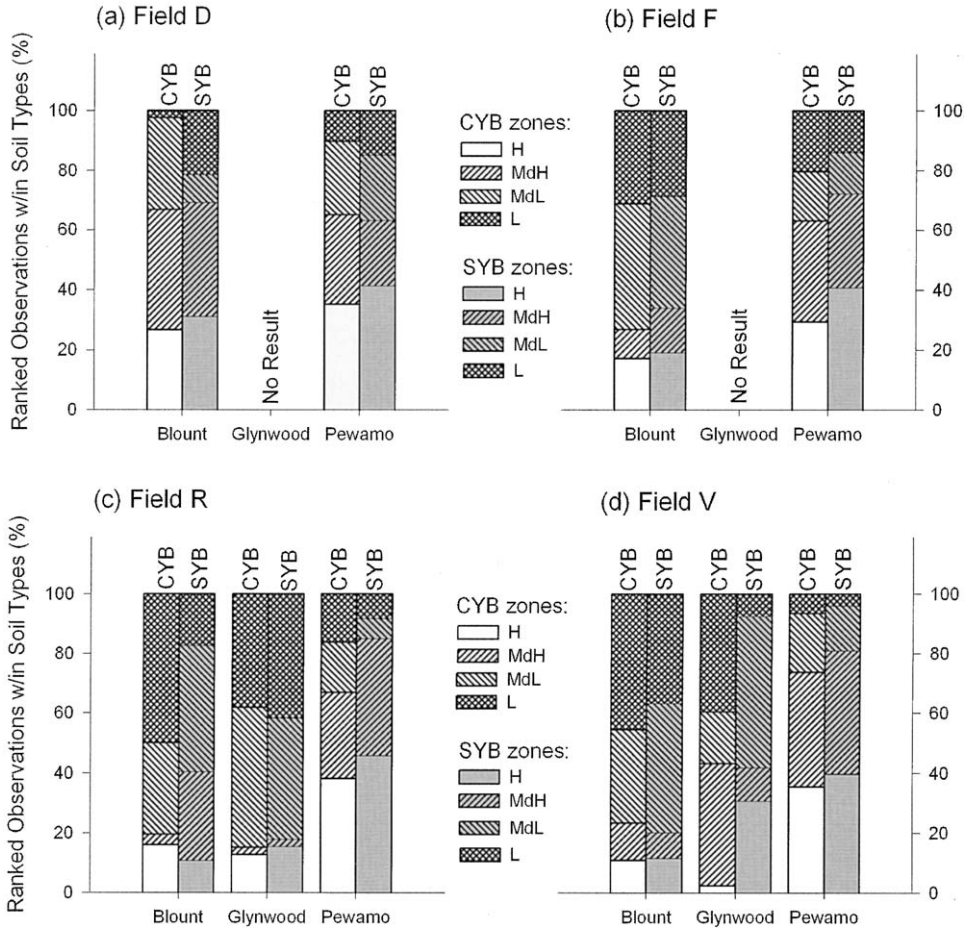

Fig. 7. For four management zone (MZ) delineations, distribution of corn yield-based (CYB) and soybean yield-based (SYB) MZ rankings within soil map units in a field. Rankings, based on MZ mean relative yields, were high (H), medium high (MdH), medium low (MdL), and low (L).

Table 7. Error rates from discriminant analysis to classify soil series from yield data. Results shown are for the separate calibration and validation data sets.

Data set	Soil series	Error rate			
		Field D	Field F	Field R	Field V
		%			
Calibration	Blount	15.7	27.8	28.6	25.8
	Glynwood	NR†	100.0	41.4	67.2
	Pewamo	61.2	21.4	15.6	10.8
	All	34.3	24.6	23.2	20.3
Validation	Blount	18.3	27.1	28.1	25.5
	Glynwood	NR	100.0	36.9	66.3
	Pewamo	60.7	15.1	16.9	9.5
	All	35.5	20.8	23.2	19.3

† No result. Insufficient observations to perform discriminant analysis.

low yielding CYB MZs and fewer than expected high or medium-high yielding MZs when compared with SYB MZ distribution. On Pewamo, there were more high and medium-high yielding SYB MZs when compared with CYB MZ distribution.

Discriminant analysis provided further insight into the relationship between soil map units and crop yields. Stepwise discriminant analysis found that all years of yield data from both crops contributed significantly to the discrimination of soil series in D, F, and R (results not shown). In V, 2000 corn did not contribute significantly to the discriminant function. Within a given field, linear discriminant functions developed using all significant factors found comparable error rates in F, V, and R (20.8, 23.2, and 19.3%, respectively) and higher error rates in D (35.5%) (Table 7). Discrimination error rates were lowest for Blount in D (18.3%) and for Pewamo in F, R, and V (15.1, 16.9, and 9.5%, respectively), corresponding to the dominant soil type within the field. Thus, with the exception of D, discrimination error rates were lower for the high-yielding Pewamo soil, suggesting that this soil may be more temporally consistent in its relative yields than the other soil series in the association.

Order 2 soil surveys are readily accessible and familiar to most farmers as a general management consideration. However, the appropriateness of using this information for precision agriculture has repeatedly been questioned (Mausbach et al., 1993; Atherton et al., 1999; Mueller et al., 2001; Bianchini and Mallarino, 2002; Franzen et al., 2002). In our study, the Order 2 soil survey suggested there would be a similar level of influence of soil series on yields among the experimental fields but the association we observed was inconsistent among fields for both crop species. Visual observation suggests that soil series in R and V may have been more distinct than in D and F as there was less elevation range in the latter two fields. Indeed, more pronounced topographical features resulting in greater spatial variability in soil water content during the growing season may explain why CYB and SYB MZs were more similar in R and V compared with D and F. Regardless, the Order 2 soil survey alone appeared insufficient to reliably identify crop productivity MZs.

SUMMARY AND CONCLUSIONS

Application of unsupervised clustering to multiple years of yield monitor data successfully identified areas within fields with different productivities. Our results suggest that cluster analysis can identify regions of common productivity, even when correlation analysis among individual years of data did not identify strong temporal stability in yield patterns. Although the soil series within study fields were expected to have marked differences in productivity potential, partitioning fields by crop-specific yield monitor data resulted in greater yield variance reduction than partitioning by either major soil series or combined productivities of both crops in the rotation.

However, this study also clearly illustrated that routine use of yield monitor data to variably manage inputs requires additional research and development of the decision rules, the software, and the underlying concepts. In our study, the NCE and FPI clustering performance measures given by the prototype, commercial-use software were not helpful in determining the optimal clustering order as they provided contradictory information. Since within MZ variance reduction in productivity is a desired outcome of partitioning fields into smaller MZs, automating the calculation of the variance reduction measures used in this study would be desirable for commercial-use software. Regardless, more scientifically founded guidance on how to interpret clustering performance measures is needed for the successful, routine application of unsupervised clustering to any agricultural data layer. Furthermore, assessing the minimum number of years of data required to develop such productivity-based MZs was beyond the scope of this study but represents a critical knowledge gap for farmers interested in PA.

The ultimate utility of productivity-based MZs was not addressed in this study and remains one of the most important, open questions in precision agriculture. Simply defining regions of differential productivity as well as identifying the existence of species-specific differences in zone productivity as we did in this study does not make this information agronomically or economically useful for management. The concept that productivity based MZs should be useful for inputs such as fertilizers can be easily traced to existing management recommendations that directly link input quantities such as fertilizers to expected plant productivity and, in the case of fertilizers, nutrient use and nutrient removal. However, to date insufficient data exist to demonstrate if and when productivity based MZs, either crop-specific or generalized for overall rotation productivity, will result in a significantly better return on input investment as compared to whole field management.

REFERENCES

Atherton, B.C., M.T. Morgan, S.A. Shearer, T.S. Stombaugh, and A.D. Ward. 1999. Site-specific farming: A perspective on information needs, benefits and limitations. J. Soil Water Conserv. 51: 66–71.

Afifi, A.A., and V. Clark. 1984. Computer-aided multivariate analysis. Lifetime Learning Publ., Belmont, CA.

Bezdek, J.C. 1981. Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York.
Bianchini, A.A., and A.P. Mallarino. 2002. Soil-sampling alternatives and variable-rate liming for soybean–corn rotation. Agron. J. 94:1355–1366.
Boydell, B., and A.B. McBratney. 2002. Identifying potential within-field management zones from cotton-yield estimates. Precis. Agric. 3:9–23.
Burrough, P.A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. J. Soil Sci. 40:477–492.
Devore, J.L. 2000. Probability and statistics for engineering and the sciences. 5th ed. Duxbury, Pacific Grove, CA.
Environmental Systems Research Institute. 1999. ArcView GIS Release 3.2. ESRI, Redlands, CA.
Evans, R.O., and N.R. Fausey. 1999. Effects on inadequate drainage on crop growth and yield. p. 13–54. *In* R.W. Skaggs and J. van Schilfgaarde (ed.) Agricultural drainage. Agron. Monogr. 38. ASA, Madison, WI.
Fraisse, C.W., K.A. Sudduth, and N.R. Kitchen. 2001. Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity. Trans. ASAE 44:155–166.
Franzen, D.W., D.H. Hopkins, M.D. Sweeney, M.K. Ulmer, and A.D. Halvorson. 2002. Evaluation of soil survey scale for zone development of site-specific nitrogen management. Agron. J. 94:381–389.
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.
Fridgen, J.J., N.R. Kitchen, K.A. Sudduth, S.T. Drummond, W.J. Wiebold, and C.W. Fraisse. 2004. Management Zone Analyst (MZA): Software for subfield management zone delineation. Agron. J. 96: 100–108.
Jaynes, D.B., and T.S. Colvin. 1997. Spatiotemporal variability of corn and soybean yield. Agron. J. 89:30–37.
Jaynes, D.B., T.C. Kaspar, T.S. Colvin, and D.E. James. 2003. Cluster analysis of spatiotemporal yield patterns in an Iowa field. Agron. J. 95:574–586.
Lamb, J.A., R.H. Dowdy, J.L. Anderson, and G.W. Rehm. 1997. Spatial and temporal stability of corn grain yields. J. Prod. Agric. 10:410–414.
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.
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. ASA, CSSA, and SSSA, Madison, WI.
McBratney, A.B., and A.W. Moore. 1985. Application of fuzzy sets to climatic classification. Agric. For. Meteorol. 35:165–185.
Mueller, T.G., F.J. Pierce, O. Schabenberger, and D.D. Warncke. 2001. Map quality for site-specific fertility management. Soil Sci. Soc. Am. J. 65:1547–1558.
Neely, T. 1987. Soil survey of Randolph County, Indiana. USDA-SCS in cooperation with Purdue University Agric. Exp. Stn. and Indiana Dep. of Natural Resources, Soil and Water Conservation Committee. USDA, Washington, DC.
SAS Institute. 2000. SAS OnlineDoc. Version 8. SAS Inst., Cary, NC.
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.
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.) Proc. Int. Conf. on Precision Agric., 4th, St. Paul, MN. 19–22 July 1998. ASA, CSSA, and SSSA, Madison, WI.
Stokes, M.E., C.S. Davis, and G.G. Koch. 2000. Categorical data analysis using the SAS System. 2nd ed. SAS Inst., Cary, NC.
Vitosh, M.L., J.W. Johnson, and D.B. Mengel. (ed.) 1996. Tri-state fertilizer recommendations for corn, soybeans, wheat, and alfalfa. Ext. Bull. E-2567. Michigan State Univ. Ext., East Lansing.