Delineating Site-Specific Management Zones and Evaluating Soil Water Temporal Dynamics in a Farmer’s Field in Kentucky

Javier Reyes,* Ole Wendroth, Christopher Matocha, and Junfeng Zhu

Due to spatial variability of soil genesis, topography, and resulting soil properties in farmers’ fields, soil and crop processes vary in space and time. Therefore, optimum rates and timing of resource applications, such as nutrients and irrigation water, may vary as well. It remains a challenge to quantify the spatial variability of a field and to identify effective ways to manage fields in a site-specific manner. The objective of this study was to delineate management zones within a farmer’s field based on relatively easily obtainable information that is statistically integrated. Moreover, soil water temporal dynamics should be evaluated regarding their spatial differences in different zones. The set of direct and indirect observations included clay and silt content, apparent electrical conductivity, soil chemical properties (pH; organic matter; and total N, P, K, Ca, Mg, and Zn), satellite-based normalized difference vegetation index (NDVI), and lidar-based topographic variables in a western Kentucky field. Several key variables and their capability to describe spatial crop yield variability were identified by using principal component analysis: soil clay content, slope, soil organic matter content, topographic wetness index, and NDVI. Two types of cluster analysis were applied to delineate management zones. The cluster analyses revealed that two to three zones was the optimal number of classes based on different criteria. Delineated zones were evaluated and revealed significant differences in corn (Zea mays L.) yield and temporally different soil moisture dynamics. The results demonstrate the ability of the proposed procedure to delineate a farmer’s field into zones based on spatially varying soil and crop properties that should be considered for irrigation management.

Abbreviations: DEM, digital elevation model; EC, electrical conductivity; ECa, apparent electrical conductivity; ECad, apparent electrical conductivity deeper depth; ECas, apparent electrical conductivity shallow depth; NDVI, normalized difference vegetation index; OM, organic matter; PCA, Principal component analysis; SOM, soil organic matter; TWI, topographic wetness index.

Understanding and managing soil and crop yield variability remains a long-standing challenge. In agroecosystems and natural systems, soil properties such as clay content, pH, soil organic matter (SOM) content, nutrient levels, and profile depth can vary drastically even within the same field (Beckett and Webster, 1971; Downes and Beckwith, 1951; Koestel et al., 2013). Precision agriculture is an approach in agroecosystem management to distribute resources site-specifically according to this variability and the associated varying input demands. The foundation of this concept is the spatial and temporal characterization of soil and crop processes through field measurements taken directly or remotely for maximizing local yield while minimizing environmental risk. Most of the research on management zone delineation has been focused on fertilization, particularly on nitrogen (N) application (Kitchen et al., 2003; Peralta et al., 2015; Ruffo et al., 2006). In other studies, weeds (Peña et al., 2013) and irrigation (Haghverdi et al., 2016; Landrum et al., 2014; Yari et al., 2017) were managed in a site-specific manner.

The application of precision agriculture has increased with advances in remote (largely distanced from the object by using platforms such as towers, vehicles, aircrafts, or satellites) and proximal (in direct contact with the object or close to it) sensing tools (Vereecken et al., 2016). Georeferenced data describing spatial and temporal variability of soil and crop state
variables and related processes can be obtained for farmers’ fields at high resolution. Yield maps, digital elevation models (DEMs), maps of normalized difference vegetation index (NDVI), or other canopy reflectance indices and apparent electrical conductivity (ECa) are among the most frequently used sources for site-specific management decisions. Yield maps are obtained from yield-monitoring systems installed on combines (Schepers et al., 2004). Digital elevation models can be derived from different sources, including existing soil maps, and at a finer resolution using lidar (James et al., 2007). Many examples exist for using ECa or electrical resistivity to define management zones based on their relationship with other important state variables, such as soil clay content (Corwin and Lesch, 2003; Moral et al., 2010; Schepers et al., 2004; Sudduth et al., 2003). The NDVI can be obtained from different sources, such as proximal sensing (e.g., Greenseeker [Walsh et al., 2012]) or remote sensing (e.g., LANDSAT, MODIS [Brown et al., 2006]). The NDVI is a canopy reflectance index that is strongly related to crop status and fitness. Because it reflects N demand over substantial parts of the growing period of many agricultural crops, NDVI has been frequently studied as a tool in site-specific N application decisions (Raan et al., 2002). Another example of remotely sensed information is the use of cosmic ray neutron probes to characterize soil moisture that has increased in recent years. Cosmic ray neutron probes measure fast neutron intensity, which strongly depends on hydrogen. This intensity can be associated with soil moisture, although it should be calibrated considering that sources of hydrogen besides soil water exist in the field (Desilets et al., 2010; Ochsner et al., 2013). Stationary probes integrate soil moisture over an area that is 100 m in diameter, getting continuous readings of temporal variation. The more recent mobile devices (cosmic ray neutron probe rovers) can cover larger areas (Finkenbiner et al., 2018; Franz et al., 2015) and can be combined with other methods, such as ECa (Gibson and Franz, 2018), although they require surveys to be performed at different times to characterize the temporal variability. Informative data can be obtained through remote and proximal sensing approaches in a much cheaper way and with higher spatial (and in some cases temporal) resolution than with collecting soil and plant samples in cumbersome field campaigns at several times during the growing season. It remains unclear, however, what soil and crop information obtained through remote or proximal sensing, including the previous years’ yield maps, is helpful to understand present-year spatial variability of soil and crop stand and to manage the field site-specifically in accordance with previous year information layers.

The challenge persists to manage field soils site-specifically to maximize biomass production efficiency and environmental benefits. Dividing a field into management zones is a promising strategy to overcome this challenge. Management zones are delineated by separating the field into different areas. Some of the areas have different response behaviors, while others may show the same behaviors (Kitchen et al., 2005). Whether areas can be considered to have homogeneous characteristics depends on the situation and is not well known. Whether or not an area is considered homogeneous depends on the variable selected. For example, the delineation of management zones can be based on crop yield maps. The spatial variability of crop yield has been reported to be related to variables such as SOM content (Mann et al., 2002), clay content (Tremblay et al., 2012), and NDVI (Teal et al., 2006). However, spatial yield patterns vary among different years because different processes during the growing season influence them (Schepers et al., 2004). Especially different weather conditions in different years can cause different spatial yield variability patterns even for the same crop species growing in the same field. The key processes and their spatial effect may vary by season, making the spatial biomass production and yield difficult to predict between different seasons. Electrical conductivity (EC) also varies in space and time, being strongly affected by soil moisture and by the salinity of sodic soils, although EC can be used to predict other variables when a strong relationship exists (Corwin and Scudiero, 2016). Moreover, EC data are spatially structured and can be combined in co-regionalization with other variables that remain stable in time, such as topography, soil depth, and clay content.

The delineation of management zones for precision agriculture based on cluster analysis has been proven to be effective to combine the impacts of different variables on the outcome (Cohen et al., 2013; Johnson et al., 2003; Li et al., 2007; Peralta et al., 2013; Vitharana et al., 2008; Yari et al., 2017). The analysis is centered on finding dissimilarities between observations by using a clustering algorithm through partitioning or hierarchical methods (Kaufman and Rousseeuw, 1990). These dissimilarities can be caused by different response behaviors between a target variable and various underlying processes. In a partitioning method, k clusters (data organized in groups) are constructed, and data are classified into k groups. Based on a selected index, the optimal number of clusters within a particular domain can be identified. For example, to work on site-specific irrigation management (e.g., Sadler et al., 2005), the right amount of water should be applied at the right time, but locations and their specific behavior in the field should also be considered. Spatial differences in topography and soil physical and chemical properties can be found within the same field. Thus, water infiltration and soil water movement also vary spatially when irrigation is applied. Examples found in the literature (Nielsen et al., 1973; Wendroth et al., 1999) illustrate the spatial and temporal variability of soil water at the field scale. Considering the spatial and temporal variability of soil water at field scale, it may be environmentally and agronomically advantageous to supply irrigation water at variable rates according to field soil water characteristics and the resulting temporal soil water dynamics. Therefore, variables that influence or correlate with soil hydraulic properties and soil water status and dynamics have to be considered in the delineation of areas. Dissimilar soil water temporal variation scales can be expected for different soil properties in different areas of the field. To analyze time-variable behavior at different zones, a wavelet analysis (Grinsted et al., 2004) is an effective strategy because it decomposes time series data in frequency and time, simultaneously allowing to observe periodic variations at different scales and times. In addition, wavelet coherence analyzes and identifies the correlation of pairs of time series data at different time scales. Studies of spatial and temporal changes in soil water using a wavelet analysis are presented by Biswas (2014), Biswas and Si (2011), and Yang et al. (2016).
The challenge of using numerical solutions to delineate management zones is to provide results that are appropriate to be used under farm conditions. Regarding the variable selection, it is essential to consider whether a specific variable represents the field variation of essential processes that underlies site-specific management or what other indirect variable could provide similar information for site-specific management, while its collection is more affordable than another more directly related variable that may be cumbersome to measure. The objective of this study was to apply easily obtainable data using proximal and remote sensing tools to define management zones in a farmer's field located in western Kentucky, which is a typical crop production region in the southeastern United States. Variables should be identified based on different approaches, and zones should be delineated by using fuzzy and hard clustering algorithms. Different approaches should be examined to evaluate if they would result in different delineations. A second objective was to evaluate differences or similarities in process behavior among delineated management zones by comparing spatial differences in corn yield and in temporal dynamics of soil moisture.

Materials and Methods

Site Description and Data Sampling

The study was conducted at Hillview Farms in Princeton, KY, Caldwell County (37°1’58.02” N, 87°51’33.06”W, 142 m asl). According to the Köppen system, the climate is classified as humid subtropical. Annual precipitation is around 1312 mm (US Climate Data, 2017), and annual mean temperature is 15°C. The maximum and minimum mean temperatures occur in June and January, respectively.

The soil in this field belongs mainly to the Crider series (Typic Paleudalfs), although some areas are classified as Nolin series (Dystric Fluventic Eutruderts) (Soil Survey Staff, 1999). In both cases, the soil texture is classified as silt loam in the surface layer. The experiment covers an area of ~27 ha. The field was cultivated with corn during the 2014, 2015, and 2017 seasons.

Soil texture at 0 to 20 cm and 20 to 40 cm depths and chemical properties at 0 to 15 cm depth were sampled across a grid of 96 points with a regular distance of 50 m (Fig. 1). Sand (0.05–2 mm) was separated by sieving. Medium (0.005–0.020 mm) and fine silt (0.002–0.005 mm) and clay (<0.002 mm) were measured with the pipette method (Gee and Or, 2002), and coarse silt (0.020–0.050 mm) was calculated as residual. Soil organic matter and total N were determined by LECO combustion. Extractable P, K, Ca, Mg, and Zn were measured by Mehlich III extraction. For pH determination, a glass electrode was used in 1:1 soil/water and Sikora buffer for Buffer pH (Jones, 2000). Apparent electrical conductivity was measured in April 2015 using a Veris 3150 (Veris Technologies, 2017). The electrodes are configured with a Wenner array, and measurements that integrate EC over a shallow...
depth (~0–30 cm) and a deep depth (~0–90 cm) were obtained, defined as EC_{as} and EC_{ad}, respectively. More details about the EC_{a} measurement are provided in Reyes et al. (2018).

A DEM was obtained at 1.5 m resolution from lidar, provided by the Kentucky Division of Geographic Information. From the DEM, slope and two widely used topographic indices (i.e., the topographic wetness index [TWI] and the stream power index [SPI]; Moore et al. [1993]) were derived. These two indices were computed by using SAGA GIS (Conrad et al., 2015). The following equations were used to obtain TWI and SPI:

$$\text{TWI} = \ln \frac{A_i}{\tan \beta}$$  \hspace{1cm} [1]

$$\text{SPI} = A_i \tan \beta$$  \hspace{1cm} [2]

where $A_i$ is the contributing catchment area (m^2 m^{-1}), and $\beta$ represents the steepest slope angle (°). The spatial pattern of TWI depicts areas that show local water accumulation and therefore high soil water saturation, and SPI indicates the potential erosional power as a result of the combined effect of slope and upstream flow.

The NDVI was obtained from Landsat 8 Operational Land Imager (Level 2 product [SR]) (USGS, 2016) and was collected during the growing seasons of corn in 2014, 2015, and 2017. The NDVI was derived from

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}$$  \hspace{1cm} [3]

where NIR is the near-infrared band (0.85–0.88 mm), and RED represents the red band (0.64–0.67 mm). High NDVI values are associated with high greenness of vegetation. Therefore, NDVI indicates crop vigor and crop N status (Brown et al., 2006). Landsat 8 images are available at a spatial resolution of 30 by 30 m. Based on the image quality and representativeness of different growth stages, images from May 2014, August 2014, and June 2015 were selected.

Management Zone Delineation

Statistical analysis was conducted in the R environment (R Core Team, 2017), and the maps for each variable were produced using raster (Hijmans, 2016) and Lattice packages (Sarkar, 2008). The analysis included measures of central tendency and dispersion and Pearson correlation using the corrplot package (Wei and Simko, 2016). Kriging and cokriging analysis as methods of interpolation were implemented to create maps by using gstat (Pebesma, 2004) and geoR (Ribeiro and Diggle, 2016) packages. Variables were log-transformed if necessary or underwent Box-Cox transformation. Semivariogram and cross semivariogram models were fitted and applied for the interpolation. For obtaining maps at a fine resolution in a reasonable computing time, all variables were interpolated in a 4-m-grid. For soil texture, a cokriging analysis was previously performed by combining soil clay content with EC_{as} (Reyes et al., 2018). Detailed steps of kriging and cokriging procedures are explained in Nielsen and Wendroth (2003).

Principal component analysis (PCA) using the FactoMineR package (Lê et al., 2008) was implemented to group the variables into statistical factors and to select key variables that explained variance in different dimensions. The sample size corresponds to the 96 soil sampling points and the intersected NDVI and topographic attributes for the corresponding locations. Principal component analysis is recommended to reduce the dimensionality of multivariate data and to work with highly correlated data (Husson et al., 2010). The Bartlett test of sphericity and the Kaiser–Meyer–Olkin test were performed to verify the adequacy of the data to be used in PCA. The obtained values were $P < 0.01$ for the Bartlett test of sphericity and 0.75 for the Kaiser–Meyer–Olkin test, suggesting that these data were applicable in PCA.

To delineate management zones, two cluster analysis methods using the cluster package (Maechler et al., 2017) were selected: a fuzzy analysis clustering (FANNY) and a hard cluster algorithm (Clustering Large Applications [CLARA]). Fuzzy cluster algorithms are preferred and are commonly used to define management zones for precision agriculture (Boluwade et al., 2015; Yari et al., 2017). The difference between the hard cluster and the fuzzy cluster is that the data are entirely assigned to one cluster in the hard cluster analysis; consequently, it is not possible to observe degrees of membership as in the fuzzy analysis. FANNY is an unsupervised soft clustering method focused on reducing the objective function, defined as

$$\sum_{i=1}^{k} \sum_{j=1}^{n(i,j)} \left( u_{i}^{r} - u_{j}^{r} \right) d(i,j)$$  \hspace{1cm} [4]

where $n$ is the number of observations, $k$ is the number of clusters, $r$ is the membership exponent, and $d(i,j)$ is the dissimilarity between the $i$th and $j$th observations.

Each data point has a membership coefficient between 0 and 1 for each cluster and is assigned to the cluster with the highest membership coefficients. A membership exponent of 2 could induce a complete fuzziness, whereas values close to 1 reduce the fuzziness. For this study, a conventional value of 1.35 (Odch et al., 1992) was used with Euclidean distance. CLARA is a hard clustering method. It is based on the same algorithm as the partitioning around medoid method described in Kaufman and Rousseuw (1990). However, in CLARA the data are divided in subsamples of fixed size. Then each subsample is partitioned in $k$ clusters.

To apply the cluster analysis, the attributes to be included have to be selected. For this purpose, all variables were standardized, and three approaches were selected to be applied in the interpolated maps:

1. Using maps of retained principal component factor scores. These factors are obtained by using the component score coefficient matrix and standardized variables (Yao et al., 2014). The retained components were selected according to the percentage of variance explained.

2. Selecting a key variable of each of the retained principal components. The selection was based on the factor loading scores obtained for each variable.
3. Combining variables from field measurements with easily obtainable remote sensing data. The selection of these variables was based on factor loadings with high contribution in the first component.

The silhouette width (Kauffman and Rousseau, 1990) was used to evaluate the most appropriate number of clusters. This method was chosen because it can be applied to both hard and soft clustering and is defined as:

\[
s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}
\]

where \(a(i)\) is the average dissimilarity between the \(i\)th point and all other points of the cluster to which \(i\) belongs. For all other clusters \(C, d(i, C)\) is the average dissimilarity of \(i\) to all observations of \(C\); then \(b(i)\) is the minimum of \(d(i, C)\), which can be seen as the dissimilarity between \(i\) and its “neighbor” cluster. The individual silhouette and the average value for each cluster is obtained. The values are in the range between -1 and 1. Negative values indicate that the data were assigned to the wrong cluster, and values closer to 1 indicate very well clustered data. Consequently, the best number of divisions will be the number of clusters with the largest average silhouette width.

**Evaluation of Management Zones**

To study the representativity of the management zones obtained, dissimilarities between clusters were quantified by observing two variables: crop yield and temporal dynamics of soil water tension. To evaluate yield differences among delineated areas, corn grain yield from the years 2014, 2015, and 2017 were used. We performed ANOVA to observe significant differences \((p < 0.05)\) in corn yield and Tukey HDS test for post hoc multiple comparison.

Temporal dynamics of soil water tension at different locations were observed as well. These measurements were made during the spring of 2016 using watermarks (Fisher and Gould, 2012) connected to antennas at four locations with a distance of 70 m for each delineated area. Data were collected each hour, and sensors were located at depths of 20, 40, and 60 cm. The method of analysis was the wavelet transform (e.g., Wendroth et al., 2011) by using the biwavelet package (Gouhier et al., 2016). The wavelet transform is a technique that can be used to analyze time series to identify cyclic variations at different frequencies or scales. This analysis includes a continuous wavelet transform for individual variables and a bivariate analysis by performing a wavelet coherence analysis (Grinsted et al., 2004; Yang et al., 2016). By using the continuous wavelet transform, we quantify the soil water tension variance for different time scales. The wavelet coherence reveals whether there exists a correlation and phase lag in the soil water tension temporal dynamics observed at depths for the same spatial zone. These analyses answer whether correlation and phase lag exist for the same depth but within different management zones. If a similar wavelet response is found for different management zones, soil water status dynamics are similar as well in response to rainfall, evapotranspiration, and soil hydraulic properties. However, for strongly differing soil properties, a different wavelet response has to be expected. The data had to be transformed with a wavelet function. In this study a Morlet wavelet was chosen as mother wavelet, which is defined as

\[
\psi_\eta(t) = \frac{\pi^{1/4}}{\nu_0} \exp\left(-\frac{\nu_0^2 t^2}{2}\right)
\]

where \(\nu_0\) is the dimensionless frequency, and \(\eta\) is the dimensionless time. The Morlet wavelet was selected because it provides a good resolution for scale and frequency. When \(\nu_0 = 6\), the Fourier period is almost identical to the scale. The continuous wavelet transforms with uniform time steps \((\delta t)\) at different scales \((s)\) of a data series \(x_n\) is defined as

\[
W_n^X(s) = \sqrt{\frac{\delta t}{s}} \sum_{n=1}^{N} x_n \psi_0\left(\frac{n' - n}{s}\right)
\]

The wavelet coherence can be considered as a correlation coefficient for common frequencies and locations. It is defined as

\[
R_n^2(s) = \left|\frac{S^{-1}(s)W_n^{XY}(s)}{\sqrt{S^{-1}(s)W_n^X(s)^2} \sqrt{S^{-1}(s)W_n^Y(s)^2}}\right|^2
\]

where \(S\) is a smoothing operator, and \(W_n^{XY}\) represents the cross-wavelet of the individual wavelet transforms \(W_n^X(s)\) and \(W_n^Y(s)\) calculated as \(W^{XY} = W^X*W^Y\), where the asterisk denotes complex conjugation.

A significance test \((p < 0.05)\) was performed, contrasting the null hypothesis that the signal is generated by a stationary process with a given background power spectrum (Grinsted et al., 2004). In the biwavelet package, this test is available only for the Morlet function.

**Results and Discussion**

The descriptive statistics of all measured variables are summarized in Table 1. In general, the average values for soil properties are similar to what is expected for fields in Kentucky (Karathanasis, 1987; Landrum et al., 2014). The silt fraction is predominant at both soil depths, but an increase in clay content is observed in the second layer. Some variables present a high coefficient of variation, such as ECs, and other chemical properties (P, K, Mg). On the other hand, the variables derived from elevation data (slope, TWI, SPI) also show a high coefficient of variation. The Pearson correlation (Fig. 2) displays some significant relationships \((p < 0.01)\): organic matter (OM) and N show the highest correlation. Clay content at both depths is strongly correlated with ECs, ECsat, slope, and Mg. The NDVI obtained in June has significant correlations with soil properties, especially silt content in the surface layer.

**Geostatistical Interpolation**

Maps for each variable were interpolated using ordinary kriging or cokriging in a 4-m grid. In Fig. 3, maps of spatial distributions for selected variables are depicted: clay content at 0 to 20 cm, clay content at 20 to 40 cm, NDVI June 2015, SOM, total soil N,
extractable P, slope, TWI, and EC<sub>as</sub>. Each of these variables reveals high spatial variability across the field. The highest clay content values are observed in the northwestern part of the field. This area also shows low NDVI values and a high slope percentage, similar to results reported by Odeh and McBratney (2000). On the other hand, the patterns observed between OM and total N content are analogous but differ from those of other variables (Fig. 3). Both OM and N present the lowest values in the southeast and some parts of the northwest. A possible explanation for these differences could be the fact that in previous years the farmer had planted Burley tobacco (Nicotiana tabacum L.) in the southeast (this area can be observed in Fig. 1), which corresponds to the area with the highest extractable P content. Tobacco has high N demand and requires more intensive management than other field crops (MacKown et al., 2000). On the other hand, TWI also reveals a high spatial variability, although with a pattern different from those of other variables. In general, only small parts of the field show high TWI values, implying the potential to have saturation overland flow (Quinn et al., 1991); therefore, runoff can occur when there is substantial precipitation or excessive irrigation. On the other hand, areas with low TWI values will potentially dry up first.

### Principal Components Analysis

After performing PCA, the three first components were retained; these components explain 65% of the total variance. The factor loadings for the variables and the three retained components are displayed in Table 2. Figure 4 presents the factor loadings of the first three components. The factor loadings most associated with the first component, which explains 40% of the total variance, were clay and silt contents at both depths, apparent electrical conductivity at both depths, and slope. Organic matter and total N strongly contributed to the second component, which explained 15% of the total variance. In the third component, representing 10% of the total variance, the highest contributions came from indices related to topography (TWI and SPI). In Fig. 4, clay content is grouped with slope. Mg content, and EC<sub>as</sub>. Silt content is grouped with NDVI measured in June 2015 and with P content. These results are consistent with the maps presented in Fig. 3, where high spatial variability for all variables was evident; however, the variables reveal different patterns. Contributions of the same variable to particular processes and their relationships vary for different fields and soil conditions. For example, the highest factor loadings for the first component were reported for different variables in other studies: pH, slope (Vitharana et al., 2008), and EC<sub>as</sub> (Van Meirvenne et al., 2013) in northwestern Europe; elevation and soil depth in Argentina (Peralta et al., 2015); and organic C and available N (Li et al., 2007) in eastern China. Furthermore, the relationships among variables varied between those studies.

### Potential Management Zone Delineation

A cluster analysis was established using the criteria described above: Criterion a, based on PCA maps of retained components (principal component [PC]1, PC2, and PC3); Criterion b, based...
on one key variable for each retained component, where clay at 0 to 20 cm, OM, and TWI were selected; and Criterion c, selecting relevant variables for the first component. In Criterion c, clay content at two depths in combination with NDVI in June 2015 and slope were chosen. In Fig. 5, the average silhouette width for a different number of clusters by using FANNY and CLARA is displayed. In both cases, results are similar, and the optimal number of divisions is 2 when the analysis is based on Criteria a and b. On the other hand, three divisions are optimal for Criterion c. According to Kaufman and Rousseeuw (1990), the results found for maximum average silhouette width can be considered as reasonably structured clusters (in the range of 0.5 to 0.7). Therefore, the results presented here can be considered appropriate to provide a well-structured division of the field. The resulting cluster zone maps are presented in Fig. 6. Similarities between the maps based on Criterion a (Fig. 6a and 6b) and Criterion b (Fig. 6c and 6d) by using both cluster methods are obvious. Criterion c (Fig. 6e and 6f) presents a zone in the northwestern part (Area 3), which has resemblance with Area 2 observed for Criteria a and b. The additional division resulting from Criterion c can be considered as a transitional zone in the middle part of the field. Nevertheless, all the criteria and methods have the same result with respect to the northwestern area, which exhibits strong dissimilarities with the rest of the field. As the farmers’ and our own field observations confirmed, this area is characterized as more eroded, and the soils in this area reveal higher clay content and steeper slopes than the rest of the field. Combining both effects, it is understandable that a soil with high

---

**Fig. 2.** Pearson correlation between measured variables. Colored cells represent significant correlation ($p < 0.01$). EC$_{ad}$, apparent electrical conductivity deeper depth; EC$_{as}$, apparent electrical conductivity shallow depth; NDVI, normalized difference vegetation index; OM, organic matter; SPI, stream power index; TWI, topographic wetness index.
clay content and a steep slope exhibits lower water infiltration and more runoff compared with a soil with only high slope or only high clay content.

Other studies have demonstrated the capability of using PCA maps for cluster analysis (Li et al., 2007; Yao et al., 2014). This implies that either all measured variables involved need to be used or that a preselection has to be made using another method (e.g., stepwise regression). On the other hand, the selection of one key variable for each retained component resulted in maps comparable to using all variables, so it could be more appropriate to use the concept of choosing key variables as a criterion. Nevertheless, it remains challenging to identify one key variable without a priori evaluation of different variables from a data set. The same criteria can be applied for other fields if their characteristics are comparable to the field in this study.

In farmers’ fields in this part of the southeastern United States, differences in local topography, as those observed in the field of this study, are typical and have a great influence on the

Fig. 3. Maps of spatial distribution of clay content at (a) 0 to 20 cm and (b) 20 to 40 cm. (c) Normalized difference vegetation index (NDVI) for June 2015. (d) Organic matter, (e) total N, (f) extractable P, (g) slope, (h) topographic wetness index, and (i) apparent electrical conductivity at a shallow depth (0–30 cm). Black dots represent the soil sampling grid. EC, electrical conductivity.
spatial variability of soil properties and, consequently, on the spatial variability of crop yield. In Fig. 7, a cross-section of an elevation profile from this experimental field is displayed with a division of the three management zones as based on Criterion c. Zones 1 and 2 are similar regarding slope, although Zone 2 is higher in clay content (Fig. 3a and 3b). Zone 3 has the steepest convex slope, which tends to cause frequent and more severe runoff and erosion events.

In general, many different processes occur in a field at different scales. Some of them are permanently visible, whereas others are only temporarily visible. Consequently, even if the desire is to classify the field in different zones, different variables can have some degree of pertinence to another. For this reason, fuzzy cluster analysis is preferred over the hard clustering method for data presented in a continuum. Moreover, the CLARA algorithm is a promising tool to perform a cluster analysis at the field scale with few divisions and using large data sets from high-resolution sources or generated by interpolation. By performing cluster analysis with R, memory limitations can cause a problem with larger datasets, whereas techniques with reduced memory usage such as CLARA could be more convenient.

Nevertheless, free source software to delineate management zones by using clustering analysis is available and has been presented in other studies (Boluwade et al., 2015; Li et al., 2007; Moral et al., 2010).

The delineation of management zones is based on a numerical solution that separates the field according to data dissimilarities. However, the results will not always be suitable for practical use. In Fig. 8, the result of dividing the field in two areas using the FANNY algorithm only based on EC as a variable is displayed. Apparent EC has been proven to be a useful variable for defining management zones (Fleming et al., 2004; Johnson et al., 2003; Moral et al., 2010; Peralta et al., 2013). Based on this map, it may appear difficult to apply a differentiated management in isolated

Table 2. Factor loadings of each variable for the first three principal components (PCs). In parentheses are the percentages of explained variance for each component. Chemical properties correspond to the surface layer (0–15 cm).

<table>
<thead>
<tr>
<th>Variable†</th>
<th>PC1 (40%)</th>
<th>PC2 (15%)</th>
<th>PC3 (10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM</td>
<td>−0.23</td>
<td>0.88</td>
<td>0.12</td>
</tr>
<tr>
<td>N</td>
<td>−0.28</td>
<td>0.80</td>
<td>0.17</td>
</tr>
<tr>
<td>pH</td>
<td>0.18</td>
<td>0.45</td>
<td>0.29</td>
</tr>
<tr>
<td>P</td>
<td>−0.65</td>
<td>−0.15</td>
<td>0.36</td>
</tr>
<tr>
<td>K</td>
<td>−0.33</td>
<td>0.59</td>
<td>−0.01</td>
</tr>
<tr>
<td>Ca</td>
<td>0.71</td>
<td>0.40</td>
<td>0.21</td>
</tr>
<tr>
<td>Mg</td>
<td>0.84</td>
<td>0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>Zn</td>
<td>−0.67</td>
<td>0.45</td>
<td>0.21</td>
</tr>
<tr>
<td>Clay 0–20 cm</td>
<td>0.89</td>
<td>0.18</td>
<td>−0.01</td>
</tr>
<tr>
<td>Silt 0–20 cm</td>
<td>−0.90</td>
<td>−0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Clay 20–40 cm</td>
<td>0.72</td>
<td>−0.09</td>
<td>0.18</td>
</tr>
<tr>
<td>Silt 20–40 cm</td>
<td>−0.84</td>
<td>0.01</td>
<td>−0.14</td>
</tr>
<tr>
<td>Slope</td>
<td>0.88</td>
<td>0.04</td>
<td>0.25</td>
</tr>
<tr>
<td>EC_{as}</td>
<td>0.70</td>
<td>0.21</td>
<td>−0.11</td>
</tr>
<tr>
<td>EC_{ad}</td>
<td>0.80</td>
<td>−0.18</td>
<td>0.28</td>
</tr>
<tr>
<td>TWI</td>
<td>−0.12</td>
<td>0.07</td>
<td>−0.83</td>
</tr>
<tr>
<td>SPI</td>
<td>0.34</td>
<td>0.29</td>
<td>−0.76</td>
</tr>
<tr>
<td>NDVI</td>
<td>−0.39</td>
<td>0.52</td>
<td>0.02</td>
</tr>
<tr>
<td>June 2015</td>
<td>−0.68</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>May 2014</td>
<td>0.34</td>
<td>0.41</td>
<td>−0.31</td>
</tr>
</tbody>
</table>

†EC_{as}, apparent electrical conductivity shallow depth; EC_{ad}, apparent electrical conductivity deeper depth; NDVI, normalized difference vegetation index; OM, organic matter; SPI, stream power index; TWI, topographic wetness index.

Fig. 4. Principal component (PC) analysis map of variables for (a) PC1 vs PC2 and (b) PC1 vs PC3. EC_{as}, apparent electrical conductivity shallow depth; EC_{ad}, apparent electrical conductivity deeper depth; NDVI, normalized difference vegetation index; OM, soil organic matter content; SPI, stream power index; TWI, topographic wetness index.
sectors classified as Zone 1 or 2 because it would require highly precise variable-rate technology. However, data can be further processed. For example, point data could be spatially aggregated or combined with data from other relevant variables to perform a clustering. In this study, EC was not directly applied in the delineation procedure but as ancillary data to produce the clay content maps presented in Fig. 3.

**Evaluating Management Zones**

Table 3 provides average corn yields for 2014, 2015, and 2017 at each of the delineated zones presented in Fig. 6. In all cases, significant differences \((p < 0.05)\) were found when comparing areas. This result indicates that the yield was different for each zone even when the field was divided into three zones. As expected, the zone with low yield was the one with the highest clay content, the largest slopes, and the lowest NDVI in June. Spatial yield patterns can turn out differently caused by each year’s specific weather conditions (Eghball and Varvel, 1997; Schepers et al., 2004). Despite this fact, these results manifest consistent zonal differences across this field that could be evaluated for site-specific management.

Another aspect to consider is the temporal stability of processes occurring in the delineated areas. In other words, the delineation presented above stands only for a time span that is represented by the underlying variables of the respective delineation. Would the same variables measured at a different time have resulted in the same delineation zones? Variables such as clay content, slope, or TWI behave rather stable over time, but others, such as NDVI, could vary during a growing season and even between years. The NDVI is an indicator of crop vigor (Teal et al., 2006), and, in the case of corn in the southeastern United States, spatial differences of NDVI can appear during growth stages with a high demand of water and nutrients (June–July). Due to its nature, this variable is affected by seasonal effects, weather conditions, and soil and crop nutrient status. However, NDVI has been shown to be a useful crop yield predictor (Wendroth et al., 2003). An example of how the delineation zones change when they are based on different NDVI data sets is presented in Fig. 9, where zones were delineated using Criterion c but with replacing NDVI in June 2015 with NDVI in June 2017. Figure 9 shows some differences compared with Fig. 3, although the northwest zone still exhibits lower values. When using the FANNY algorithm to divide the field in three areas, the patterns in Fig. 9b and Fig. 6e are similar. In this example, some differences in NDVI were identified during different years, manifested as extreme values in certain areas. Nevertheless, these differing NDVI data, in addition to other rather stable variables, resulted in similar delineation zones.

Regarding soil moisture, it was expected in this study that the soil textural composition in different zones of the field affects the temporal dynamics of soil water tension. For example, in a zone with relatively low clay and high silt content, changes in tension at a given depth as a consequence of rainfall or evapotranspiration would occur at a slower rate than in a zone with a higher clay and lower silt content due to the differences in shape of the soil water retention curve and hydraulic conductivity function in the different zones. Therefore, the question addressed here was whether the three different delineated zones would reveal differences in temporal soil water tension dynamics and whether these differences would depend on the similarity or diversity between different zones. For this purpose, soil water tension at three depths during the spring of 2016 was monitored at different locations distributed across the different zones at 1-hour intervals. This variable and the experimental set-up used here are widely applied for scheduling irrigation (Liang et al., 2016). For site-specific irrigation management, it is recommended to install soil moisture sensors in each defined zone, but the minimum number varies depending on the characteristics of the field (Barker et al., 2017; Tollner et al., 1991). Due to similarities in the delineation derived from different approaches, this analysis was centered on the comparison of three areas using FANNY (Fig. 6e). In Fig. 10, the soil water tension at different depths for each area and daily precipitation are presented, where each line represents the average value of four sensors installed in each zone. Between rainfall events, soil water tension increases as a consequence of evapotranspiration but...
Fig. 6. Maps of delineated management zones based on maps of principal component (PC) factor scores by using (a) FANNY and (b) CLARA based on clay content at 0 to 20 cm, organic matter (OM) content, and topographic wetness index (TWI) by using (c) FANNY and (d) CLARA; and based on clay content at 0 to 20 and 20 to 40 cm, normalized difference vegetation index (NDVI) from June 2015, and slope by using (e) FANNY and (f) CLARA.
with different behavior in each zone. Zone 3, identified with high clay content and slope, shows the steepest tension increase during drying not only at 20 cm depth near the surface but also at 40 and 60 cm. On the other hand, during days with precipitation, soil water tension reaches lower values than at the other two areas. Values in Zone 1 fluctuate less but are more similar to Zone 2. This behavior is interpreted to be a consequence of higher silt content and higher water capacity than in the clay soil.

The temporal dynamics of soil water tension and their variability at different scales should be analyzed to detect similarities or differences in their dynamics and to understand how different temporal behavior over the time period is manifested in wavelet spectra. Results of wavelet analysis for each zone and depth are presented in Fig. 11. This analysis allows identifying cyclic patterns at different frequencies and at different points or periods in times, here represented in periods of 1 h with temporal scales that range from hours to several days. Furthermore, time-specific periodic features become apparent. Results show that temporal fluctuation behavior shows some differences among zones and depths. High spectral density values were found for a frequency equivalent to 5 to 10 d, which coincides with the occurrence of rainfall events. In addition, a significant cyclic pattern at lower frequencies during rainfall periods was identified.

The wavelet coherence between time series observed at the same location but at different depths is presented in Fig. 12. As expected, the coherence is significant \( (p < 0.05) \) at frequencies of 5 to 10 d between layers for the same zone. In Zone 3, a strong coherence for frequencies of 1 to 2 d is found. The phase represented by arrows presents mainly an in-phase behavior, which means that changes in soil water tension occur synchronously or without a substantial lag at different depths when phase lag is close to 0°. The coherence is higher when layers of 20 and 40 cm are compared and decreases when other combinations (20 vs. 60 cm or 40 vs. 60 cm) are contemplated because redistribution of rain water and removal of water through evapotranspiration occurs faster in shallow layers than at the greater depths monitored here. This finding can also be

Table 3. Average corn yield among different management zones by using different variables and cluster algorithms.

<table>
<thead>
<tr>
<th>Variables selected†</th>
<th>Cluster zones</th>
<th>Corn yield (Mg ha⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2014</td>
</tr>
<tr>
<td>PC1 + PC2 + PC3</td>
<td>1</td>
<td>12.5a</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.6b</td>
</tr>
<tr>
<td>Clay 0–20 cm + OM + TWI</td>
<td>1</td>
<td>12.5a</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>11.6b</td>
</tr>
<tr>
<td>Clay 0–20 cm + clay 20–40 cm + slope + NDVI, June 2015</td>
<td>1</td>
<td>12.6a</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12.2b</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>11.4c</td>
</tr>
</tbody>
</table>

† NDVI, normalized difference vegetation index; OM, soil organic matter; PC, principal component; TWI, topographic wetness index.
‡ Different letters indicate significant differences \( (p < 0.05) \) between cluster zones.
attributed to the fact that results at 60 cm represent the wettest soil conditions with small fluctuations (Fig. 10). Bearing in mind these relationships, a method such as the exponential filter (Albergel et al., 2008; Wagner et al., 1999) could be applied to predict the soil moisture in deeper layers based on the measurements at 20 cm depth, where we could expect to obtain a good prediction at 40 cm depth due to the stronger temporal correlation.

Wavelet coherence for soil water tension between areas is presented in Fig. 13. Significant values \( p < 0.05 \) were observed primarily when comparing Zones 1 and 2 at 20 and 40 cm. On the other hand, Zone 3 in general does not have high coherence with the other two zones. This result informs us that Clusters 1 and 2 at 20 and 40 cm depth are zones in which the soil water temporal dynamics are significantly correlated. On the other hand, the temporal
Fig. 11. Continuous wavelet transform at different depths for each cluster zone. Period resolution is 1 h. Significant differences ($p < 0.05$) against red noise are shown as a thick contour. The light shade represents the area outside of the cone of influence.
Fig. 12. Wavelet coherence between depths for individual cluster zones. Period resolution is 1 h. Significant differences ($p < 0.05$) against red noise are shown as a thick contour. The light shade represents the area outside of the cone of influence. Arrows indicate the relative phase relationship: right, in-phase; left, anti-phase; down, first series leading second series by $90^\circ$; up, second series leading first series by $90^\circ$. 
Fig. 13. Wavelet coherence between cluster zones at different depths. Period resolution is 1 h. Significant differences ($p < 0.05$) against red noise are shown as a thick contour. The light shade represents the area outside of the cone of influence. Arrows indicate the relative phase relationship: right, in-phase; left, anti-phase; down, first series leading second series by 90°; up, second series leading first series by 90°.
variation presents differences in space when we compare the other possible combinations (i.e., 1 vs. 3 and 2 vs. 3 at each depth and 1 vs. 2 at 60 cm). The phase relationship shows that arrows mainly point downward, indicating that, at the same depth, there is a lag in the process of soil water tension between zones (i.e., the first series leads the second by a phase shift of 90 degrees). This shows that temporal changes in soil water tension do not occur synchronously, although the patterns present high correlation, which is caused by the delay due to different soil hydraulic properties in the different textural zones. Yang et al. (2016) also found high coherence during rainfall events. Moreover, when comparing different zones, dissimilarities in temporal variability were observed: Zone 3 (at the northwestern part of the field) reveals the most severe changes in time. Between rainfall events, the soil dries out faster in Zone 3 than in the other zones. On the other hand, when high precipitation amounts are received, soil water tension can reach values closer to saturation within a shorter time. This result resembles what commonly occurs in clayey soils due to their water retention characteristics. Combined with the fact that this is the area with the highest slope, it is not recommended to manage Zone 3 in the same way as Zones 1 and 2 because of its limited water capacity and infiltrability. Regarding irrigation, the optimal time, frequency, and rate of water application differs among management zones in this field. For example, the clayey zones should be irrigated more frequently and with lower rates than the other zones to apply the required amount of water at a rate that would not cause runoff but that can be taken up by the soil.

Conclusions
Spatial differences in local topography are a key driver of spatial variability in the studied field. Areas with high slope were eroded, resulting in higher clay content closer to the soil surface. This might explain the lower crop productivity as reflected by NDVI and crop yield and influenced SOM and total N content in the soil.

Cluster analysis resulted in a division of two or three areas, depending on the underlying criteria. A method based on PCA maps revealed a similar delineation compared with selected key variables, which in this study were clay content, SOM content, and topographic wetness index. For larger data sets, CLARA proves to be as good as the FANNY clustering when used in fields with large data sets and a small number of divisions.

Delineated areas changed when different variables were selected but showed concordance at identifying the most restrictive zone for agricultural practices. With all the different delineation approaches, crop yields significantly differed between delineated zones. Soil water dynamics revealed cyclic patterns of 5 to 10 d and cyclic variation at smaller time scales on specific dates. In both cases, these were mainly influenced by precipitation. The behavior of temporal dynamics differs among zones, most probably due to site-specific soil hydraulic properties. The results from this study emphasize the need for delineating functional subunits in farmers’ fields, and the methodology provides a feasible way to delineate site-specific management zones to improve the productivity of this field, as shown in the various ways to evaluate the delineation. The method of dividing this field into different functional and management units obtained in this study is suitable for practical use.

Acknowledgments
This research was supported by the Kentucky Small Grain Growers’ Association; the Kentucky Corn Growers’ Association; the Kentucky Soybean Board; the Southern Soybean Research Program; the SFB 271 Water Quality Program through the College of Agriculture, Food and Environment and the Department of Plant & Soil Sciences; and the USDA National Institute of Food and Agriculture (Multistate Project KY006093). We thank Riley Jason Walton and James Dolarhie for technical support, Xi Zhang for support during the field experiments, and Trevor Gilkey (Hillview Farms, Princeton, KY) for allowing this research to be conducted on his farm. This is Publication no. 18-06-088 of the Kentucky Agricultural Experiment Station, which is published with the approval of the Director.

References
doi:10.5194/hess-12-1323-2008