2004

Using Discriminant Analysis and GIS to Delineate Subsurface Drainage Patterns

Allah Bakhsh
Iowa State University

Ramesh S. Kanwar
Iowa State University, rskanwar@iastate.edu

Using Discriminant Analysis and GIS to Delineate Subsurface Drainage Patterns

Abstract

The contamination of soil and water resources from nutrients, transported in subsurface drainage water having different drainage patterns, has important repercussions on the ecological environment and human health. This study was designed to delineate subsurface drainage patterns using cluster analysis based on six years (1993 to 1998) of field measured data on subsurface drainage flows from thirty-six 0.4 ha field experimental plots. These drainage patterns then were related spatially to the soil and topographic attributes using discriminant analysis and the map overlay capability of Geographic Information Systems (GIS) to develop cause-effect relationships. The experimental field plots, under various tillage and nitrogen management treatments, were located on glacial till derived soils at Iowa State University's Northeastern Research Center near Nashua, Iowa. The field-measured subsurface drainage volumes were normalized to make comparisons over all plots and years, and the normalized data were used in the subsequent statistical and GIS analyses. After performing cluster analysis, the output was generated as GIS data layers showing low, medium, and high drainage areas. Stepwise discriminant analysis identified elevation, slope, and average normalized yield as the factors contributing significantly ($P < 0.10$) to the formation of subsurface drainage zones. GIS data layers of the factors, identified during discriminant analysis, were overlaid on the drainage patterns to study the spatial relationships. Map overlay analysis showed that high drainage areas were consistently found at low elevation levels in the vicinity of Floyd soils over the 6-year study period. The combined use of discriminant analysis and GIS was found to be effective in delineating subsurface drainage zones so that appropriate management practices can be applied to mitigate the environmental effects resulting from medium and higher subsurface drainage effluents.

Keywords

Cluster analysis, Map overlay, Soil attributes, Topographic attributes

Disciplines

Agriculture | Bioresource and Agricultural Engineering | Water Resource Management

Comments

This article was published in Transactions of the ASAE. Vol. 47(3): 689-699, doi:10.13031/2013.16101. Posted with permission.

USING DISCRIMINANT ANALYSIS AND GIS TO DELINEATE SUBSURFACE DRAINAGE PATTERNS

A. Bakhsh, R. S. Kanwar

ABSTRACT. *The contamination of soil and water resources from nutrients, transported in subsurface drainage water having different drainage patterns, has important repercussions on the ecological environment and human health. This study was designed to delineate subsurface drainage patterns using cluster analysis based on six years (1993 to 1998) of field measured data on subsurface drainage flows from thirty−six 0.4 ha field experimental plots. These drainage patterns then were related spatially to the soil and topographic attributes using discriminant analysis and the map overlay capability of Geographic Information Systems (GIS) to develop cause−effect relationships. The experimental field plots, under various tillage and nitrogen management treatments, were located on glacial till derived soils at Iowa State University's Northeastern Research Center near Nashua, Iowa. The field−measured subsurface drainage volumes were normalized to make comparisons over all plots and years, and the normalized data were used in the subsequent statistical and GIS analyses. After performing cluster analysis, the output was generated as GIS data layers showing low, medium, and high drainage areas. Stepwise discriminant analysis identified elevation, slope, and average normalized yield as the factors contributing significantly ($P < 0.10$) to the formation of subsurface drainage zones. GIS data layers of the factors, identified during discriminant analysis, were overlaid on the drainage patterns to study the spatial relationships. Map overlay analysis showed that high drainage areas were consistently found at low elevation levels in the vicinity of Floyd soils over the 6−year study period. The combined use of discriminant analysis and GIS was found to be effective in delineating subsurface drainage zones so that appropriate management practices can be applied to mitigate the environmental effects resulting from medium and higher subsurface drainage effluents.*

Keywords. *Cluster analysis, Map overlay, Soil attributes, Topographic attributes.*

The contamination of soil and water resources with chemicals, transported from agricultural fields, has important repercussions on the ecological environment and human health. Several studies have been conducted to evaluate the degree of water contamination and the possible sources of pollution of water bodies (Kanwar and Bakhsh, 2001; Randall and Mulla, 2001; Jaynes et al., 1999). In this context, a subject that has received particular attention is the discharge of chemicals from agricultural lands through subsurface drainage water, particularly in the Midwestern parts of the U.S. Use of artificial subsurface drainage is an important and common practice used to maintain agricultural productivity of the poorly drained soils in the Midwest, where over 30% of the soils have been drained using subsurface drainage network (Kanwar et al., 1999; Hatfield et al., 1998; Randall et al., 1998).

Subsurface drainage systems remove excess water from the root zone of agricultural lands to minimize the adverse effects of excess water on crop productivity. In addition to maintaining a favorable soil water and air balance in the root zone, subsurface drainage has been found to be responsible for transporting agricultural chemicals from the bottom of the root zone to the edge of the field (Bakhsh et al., 2002). Subsurface drainage systems also short−circuit the flow of water by draining the top of the saturated zone directly into streams in the Midwest, and eventually into the Mississippi River (Goolsby et al., 2001). The role of subsurface drainage becomes more important when water drained from agricultural fields joins streams/ditches or surface water bodies used as drinking water sources. Recently, the development of a hypoxia zone in the Gulf of Mexico has been linked to nitrate leaching from agricultural lands through subsurface drainage water (Rabalais et al., 1999). Therefore, monitoring and evaluation of subsurface drainage systems on a long−term basis can help assess the impact of subsurface drainage on the transport of chemicals from agricultural fields (Kanwar et al., 1997; Randall and Mulla, 2001).

Subsurface drainage from agricultural land is a function of climate, soil, topography, and management factors. Subsurface drainage volumes can vary considerably over the years because of changing weather conditions. In addition to this temporal variability, a substantial spatial variability in subsurface drainage effluents from field to field has also been reported (Bakhsh et al., 2002). Subsurface drainage systems also contribute significantly to leaching of agrochemicals from agricultural fields. Therefore, it becomes critical to investigate the spatial and temporal patterns of subsurface drainage effluents and determine the possible cause−effect relationships for their occurrence (Bakhsh et al., 2000a).

Article was submitted for review in November 2002; approved for publication by the Soil & Water Division of ASAE in March 2004.

Use of trade name is for reader information and does not imply any endorsement by Iowa State University.

The authors are **Allah Bakhsh, ASAE Member Engineer,** Post−Doctoral Research Associate, and **Ramesh S. Kanwar, ASAE Member Engineer,** Professor and Department Chair, Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, Iowa. **Corresponding author:** A. Bakhsh, 102 Davidson Hall, Iowa State University, Ames, IA 50011; phone: 515−294−4302; fax: 515−294−2552; e−mail: bakhsh@iastate.edu.

  ISSN 0001−2351

One such approach used to identify patterns is cluster analysis, and discriminant analysis is used to determine the contribution of various factors in the formation of clusters (Young and Hammer, 2000; Li et al., 1992). Both of these approaches have been used successfully in classification of soils (Brejda et al., 2000; Al−Abed et al., 1989; Seelig et al., 1991) and other disciplines such as medical science (Norton et al., 2000), geology (Feldman et al., 1991; Jones, 1989), environmental sciences (Wu et al., 2001; Ayana and Bekele, 1999), and precision agriculture (Jaynes et al., 2003). Soil science studies have also applied these methods to classify drainage classes (Kravchenko et al., 2002) and soil classes (Brejda et al., 2000). The idea is to statistically minimize within−group variability while maximizing among−group variability in order to produce relatively homogenous groups that are distinct from one another (Young and Hammer, 2000). The term clustering is synonymous with numerical taxonomy, pattern recognition, and clumping (Mardia et al., 1979). Cluster and discriminant analyses are two multivariate statistical methods, and their use in classification is appropriate because the properties are not independent, but interrelated in a complex way. These methods are often useful for the simplification and consequent understanding of a complex soil situation (Li et al., 1992).

Kravchenko et al. (2002) applied discriminant analysis and geostatistics to create drainage maps using topographical and soil properties data. Al−Sulaimi et al. (1997) grouped drainage basins statistically and applied discriminant analysis to confirm such grouping. Ayars and Meek (1994) used cluster analysis to group individual drainage sumps and concluded that flow−load relationships were better characterized. Bari (1992) suggested a decision support system based on cluster analysis to pre−process a large number of alternatives in water resources planning and decision making. These studies show the potential of cluster and discriminant analysis to group the objects and study their patterns. The spatial occurrence of these patterns can be investigated further by overlaying data layers of various soils and landscape attributes using Geographic Information Systems (GIS) (Bakhsh et al., 2000b).

Many studies have investigated the effects of soil properties on drainage using GIS because of its capability to analyze and integrate the spatial data layers (Rafaelli et al., 2001; Corwin et al., 1999; Wang and Yin, 1998). GIS is a powerful computer software capable of capturing, storing, retrieving, analyzing, and displaying spatial data to understand their occurrence and relations with other soil−related attributes. The general approach for using GIS is to display factors important in an analysis as separate data layers and overlay them to determine the integrated effects of these factors (Bakhsh et al., 2000b). Subsurface drains integrate the effects of spatial variability on a field scale and can be better investigated using GIS map overlay capabilities. A single subsurface drainage line can possibly drain many soils within its drainage divide and can be affected by soil type and topography. Therefore, drainage patterns need to be investigated using long−term data to determine their spatial and temporal stability. This study was designed to investigate the spatiotemporal variability in drainage patterns using cluster analysis over a long−term basis and correlate the occurrence of drainage patterns with soil−related attributes such as soil type and topography. No other study, however, has been conducted to overlay the output of cluster analysis by GIS data layers of soil type, elevation, slope, and aspect surfaces to analyze their spatial relationships. Therefore, the specific objectives of the study were:

- To develop GIS data layers of subsurface drainage patterns to study the temporal and spatial variability and overlay them on the digital elevation model and soil type map to establish the cause−effect relationships.
- To delineate subsurface drainage patterns using cluster analysis based on six years (1993 to 1998) of field measured data on subsurface drainage flows from 36 field plots and relate their occurrence with the soil and landscape attributes using discriminant analysis.

Materials and Methods

Iowa State University's northeastern research center at Nashua, Iowa, has 36 experimental plots (each 0.4 ha, 58.5 × 67 m in size) underlain by subsurface drainage lines. The soils at the site are Floyd loam (fine−loamy, mixed, mesic Aquic Hapludolls), Kenyon loam (fine loamy, mixed, mesic, Typic Hapludolls), and Readlyn loam (fine loamy, mixed, mesic Aquic Hapludolls) (Karlen et al., 1991). These soils are moderately well to poorly drained and lie over glacial till. From 1978 to 1992, the site was under four tillage treatments (chisel plow, moldboard plow, ridge till, and no−till) and two crop sequences of continuous corn and corn−soybean (*Zea mays* L. and *Glycine max* L.) rotation. From 1993 to 1998, tillage treatments were reduced to two (chisel plow and no−till) by including additional nitrogen management treatments while keeping the same crop sequence at the site. More details about various experiments at the research site can be found in Bakhsh et al. (2002).

The subsurface drainage system at the site was installed in 1979 to monitor the drainage flow rate for each individual plot. Each plot has a separate drainage sump equipped with automatic flow recorders. Each plot is drained separately and has subsurface drainage lines installed in the center of the plot at a depth of 1.2 m below the ground surface with a drain spacing of 28.5 m. The cross contamination of each plot was checked by installing subsurface drainage lines on the northern and southern borders of the plot and isolating the eastern and western borders with berms (Kanwar et al., 1999). The central subsurface drainage lines are intercepted at the end of the plots and are connected to individual sumps for measuring drainage effluents and collecting water samples for chemical analysis. The sumps are equipped with a 110 V effluent pump, water flowmeter, and an orifice tube to collect water samples. Data loggers, connected to the water flowmeters, record subsurface drainage flow continuously as a function of time. Approximately 0.2% of the water pumped from the sump flowed through a 5 mm diameter polyethylene tube to a water sampling bottle located in the collection sump each time the pump operated. Cumulative subsurface drain flows were recorded, and sampling bottles were removed two times per week from mid−March to the beginning of December during the entire study period. A more detailed description of the automated subsurface drainage system installed at the site can be found in Kanwar et al. (1999).

Statistical Analysis

Descriptive statistics were calculated using the PROC UNIVARIATE procedure (SAS, 2000) to determine variance and distribution of data. Subsurface drainage data from 36 field plots were collected from 1993 to 1998 with changing rainfall conditions over the study period. The subsurface drainage volumes also changed considerably over the years as a result of change in climatic conditions. To compare data among years, and to remove the cropping systems (corn/soybean) and various nutrient treatment effects on subsurface drainage effluent, subsurface drainage data from all 36 field plots were normalized on a yearly basis. Treatment effects on subsurface drainage data were not found to be significant (Bakhsh et al., 2002). To investigate spatiotemporal variability in drainage effluents due to intrinsic factors, the extrinsic factor of rainfall having significant ($P < 0.05$) effect on the drainage was removed. Therefore, the effect of "more rain, more drain" was removed by normalizing the subsurface drainage data for each plot on yearly basis. The data failed the Kolmogrov−Smirnov test of normality and therefore was standardized using a robust approach, as described below (Bakhsh et al., 2000b):

$$z_j = \frac{y_j - y'_j}{s_j} \tag{1}$$

where z_j is normalized annual subsurface drainage volume for each plot for the jth year, y_j is the annual subsurface drainage volume of each plot for the jth year, y'_j is the median of subsurface drainage volume for the jth year, and s_j is the estimate of subsurface drainage variation for the jth year. Similar approaches have been used by Jaynes and Hunsaker (1989), and Colvin et al. (1997). Median estimates were used for y'_j (Cressie, 1993) because subsurface drainage data were not normally distributed. The interquartile range (75th to 25th percentile) was used as an estimate of s_j. As robust estimators, the median and interquartile range reduce the impact of outliers and non−normality on the calculation of z_j (Colvin et al., 1997). The normalized subsurface drainage data were used during all the statistical as well as GIS analyses for comparison over all the fields and years.

Crop yield data were included in the study to take into account the effect of evapotranspiration on subsurface drainage volumes. Corn and soybean yield data were normalized separately for each crop. Corn data were normalized for each treatment for each year and soybean yield data were normalized on yearly basis because no fertilizer was applied to soybean. After normalizing data for both the crops for six years, average yield data were obtained for each plot over six years, which were used during statistical and GIS analysis. Normalization of crop yield data was necessary to remove crop effects on drainage and over the years with changing climatic conditions (Bakhsh et al., 2000b).

GIS Analysis

ArcView (3.2) GIS software was used to digitize the soil type map using the polygon option after scanning the soil map of the study area. The ArcView Image Analysis extension was used to rectify the scanned image with the feature (plots layout theme) of the study area. The field boundary theme and study area plot layout theme were drawn following the display of elevation theme data collected during a detailed topographic survey of the study area (Singh, 1994). Elevation data along with coordinates were measured at the 96 data points following a regular grid of 76 × 29 m using automatic level (Sokkia Co. Ltd, Japan). These elevation data were used to build a digital elevation model (DEM) for the site. The ArcView Spatial Analyst extension (ESRI, 1996) was used to create an elevation surface from 96 data points. A spline kriging method with the six closest neighboring points (Kravchenko and Bullock, 2000) was used to interpolate the elevation surface. Spline interpolation is preferred because it is best suited for gently varying surfaces such as elevations, water table heights, or pollution concentrations (ESRI, 1996, p. 92).

From the DEM, slope and aspect data layers were derived for the site. Zonal functions in ArcView were used to compute an output table for average elevation, slope, and aspect data for 36 plots, which were used in stepwise discriminant analysis. Similar procedures were used to build drainage data layers using normalized drainage data of 36 plots for each year from 1993 to 1998. All the normalized drainage data layers were reclassified into five classes (<−2, −2 to −1, −1 to 1, 1 to 2, and >2) to compare drainage variability and patterns over the years. All the normalized drainage data layers were overlaid by DEM and plot layout theme to assess stability in drainage patterns over the years and their relationship with topography. The ArcView map query option was used to overlay and integrate all the six drainage data layers to determine the new theme containing drainage area >1 standard deviation (high drainage) among all these layers. This new integrated theme, showing high drainage areas, was overlaid by soil type and elevation contour themes to determine the spatial relationships.

Cluster and Discriminant Analysis

Jardine and Sibson (1971, p. 276) defined a cluster as "a set of objects characterized by properties of isolation and coherence." Gengerelli (1963) defined a cluster as "an

Table 1. Descriptive statistics for subsurface drainage data collected from 36 plots over a 6−year period (1993 to 1998) at Nashua, Iowa.

Statistic	1993	1994	1995	1996	1997	1998	Avg.[a]
	Subsurface Drainage (mm)						
Mean	387	82	137	62	100	228	166
Median	370	67	108	53	72	213	103
Standard deviation	153	62	103	42	85	115	150
Skewness	0.89	2.3	2	1.8	2	1.6	1.6
Kurtosis	1.7	4.9	4	3.4	4.4	3.7	2.7
Minimum	66	22	14	16	2.1	44	2.1
Maximum	781	287	464	185	377	588	781
Interquartile range	140	36	63	31	67	82	177
CV (%)	39	75	75	67	85	51	90
	Normalized Drainage Data						
Mean	0.12	0.43	0.47	0.32	0.42	0.18	0.32
Median	0	0	0	0	0	0	0
Standard deviation	1.1	1.7	1.6	1.4	1.3	1.4	1.4
Skewness	0.9	2.3	2.0	1.8	2	1.6	1.9
Kurtosis	1.7	4.9	4.0	3.4	4.4	3.7	4.4
Minimum	−2.2	−1.3	−1.5	−1.2	−1.0	−1.9	−2.2
Maximum	2.9	6.2	5.7	4.3	4.6	4.6	6.2
Interquartile range	1	1	1	1	1	1	1
CV (%)	892	401	352	433	306	769	442
	Growing Season (March − November)						
Rainfall (mm)	1030	750	800	680	750	980	840[b]

[a] 6−year average (1993−1998).
[b] 30−year average.

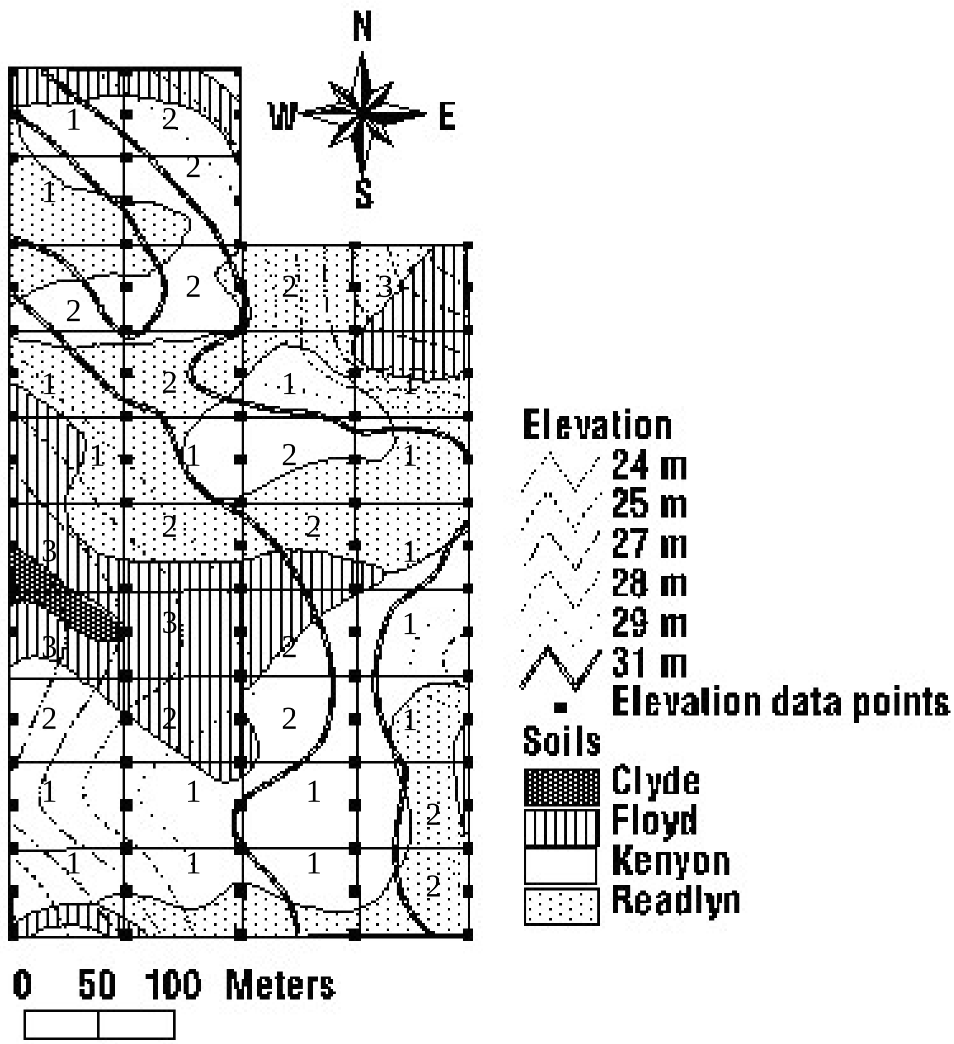

Figure 1. Soil type, topography, and 36 drainage plots of the study area at Nashua, Iowa, showing three drainage clusters (1 = low, 2 = medium, and 3 = high).

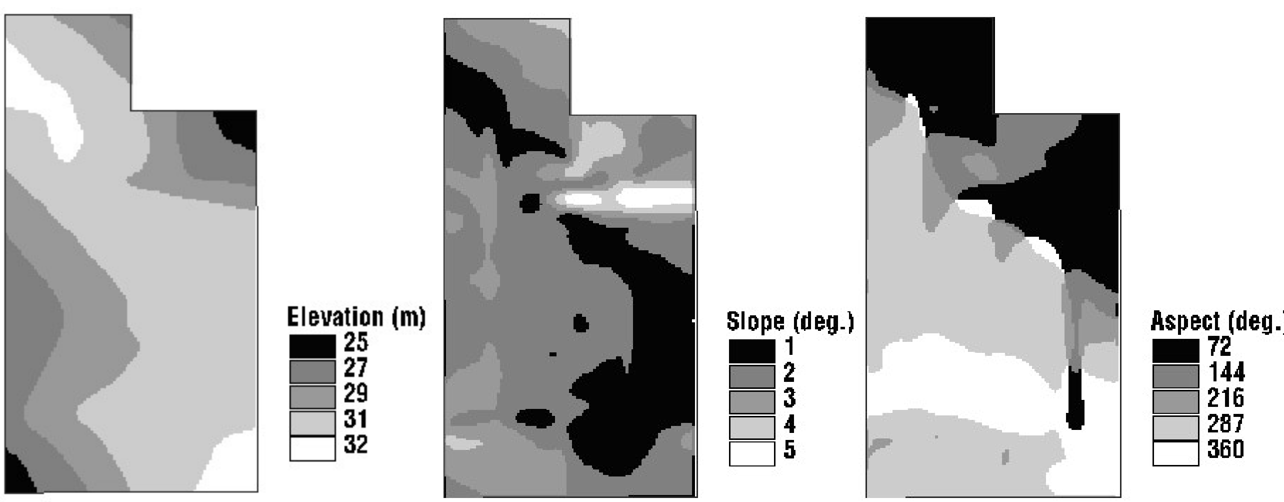

Figure 2. Topographic attributes of the study area: elevation, slope, and aspect data layers.

aggregation of points in test space such that the distance between any two points in the cluster is less than the distance between any two points inside and outside the cluster." The PROC FASTCLUS procedure (SAS, 2000) with 10 iterations and zero convergence criteria was used to develop clusters based on annual subsurface drainage data from 36 plots for six years. The FASTCLUS procedure in SAS combines an effective method for finding initial clusters with a standard iterative algorithm for minimizing the sum of squared distances from the cluster means. The result is an efficient procedure for disjoint clustering of large data sets with the option of specifying number of clusters. This procedure uses a nearest−centroid sorting method, and each observation is assigned to the nearest seed to form temporary clusters. The initial seeds are replaced by the cluster means, and the process is repeated until no further changes occur in the clusters. More details can be found in the SAS documentation (SAS, 2000). Different numbers of clusters were tried to get the best results, and finally three clusters were selected based on evaluation criteria of cluster formation (pseudo−F statistic, R^2, and cubic clustering criterion; SAS, 2000). The cluster output was plotted showing the exact location of each member in the cluster when coordinate data of each member were given in the dataset used in SAS.

After cluster formation, a stepwise discriminant procedure (PROC STEPDISC; SAS, 2000) was used to determine the contribution of various soil attributes of elevation, slope, aspect, soil, and average normalized crop yield in the formation of clusters. The STEPDISC procedure performs a stepwise discriminant analysis to select a subset of the quantitative variables for use in discriminating among the clusters. STEPDISC is similar to stepwise regression analysis and incorporates each variable into the model based on its significance level ($P < 0.15$). This procedure uses forward selection, backward elimination, or stepwise selection technique. This procedure is useful to establish cause−effect relationships for cluster occurrence and is helpful for applying management practices based on zones defined by clusters. Verification of cluster formation was made using the discriminant procedure, i.e., PROC DISCRIM (SAS, 2000), to assess how accurately the clusters can be predicted using the variables selected during the STEPDISC procedure. This procedure derives canonical variables (linear combinations of the quantitative variables) that summarize variations between clusters. More details about these procedures can be found in the SAS documentation (SAS, 2000). In addition, the cluster analysis output was used as input in ArcView to create a cluster grid to show low, medium, and high drainage areas and was overlaid by the DEM data layer. The cluster analysis output was also overlaid by the variables selected during the stepwise discriminant analysis to study the spatial relationships.

RESULTS AND DISCUSSION

SUBSURFACE DRAINAGE

Subsurface drainage was affected by the growing season rainfall amount because a significant ($P < 0.05$) correlation ($R^2 = 0.89$) was observed between rainfall and subsurface drainage volume over the study period (1993 to 1998). Average annual subsurface drainage volume varied from a low of 62 mm in 1996 with an annual rainfall of 680 mm to a high of 387 mm with an annual rainfall of 1030 mm in 1993 (table 1). This compares to a 30−year average annual rainfall of about 840 mm for the study area (Voy, 1995). The year 1993 was a wet year, having annual rainfall 23% greater than the 30−year average annual rainfall. All other years were lower than the 30−year average annual rainfall (750 mm for 1994, 800 mm for 1995, and 750 mm for 1997) except 980 mm for 1998, which was 17% more than the 30−year average annual rainfall. The 6−year average subsurface drainage volume (166 mm) showed that about 20% of the average growing season rainfall (832 mm) was drained in the form of subsurface drainage flow for the study area.

The descriptive statistics on drainage data showed that minimum subsurface drainage volume on an individual field basis ranged from 2 mm in 1997 to 66 mm in 1993, and maximum subsurface drainage volume varied from 185 mm in 1996 to 781 mm in 1993 (table 1). The coefficient of variation ranged from a low of 39% in 1993 to a high of 85% in 1997. This analysis showed that despite temporal variability in subsurface drainage volume from year to year, there was significant variability in subsurface drainage data on a field−to−field basis, which could be due to climate, soil, topography, and management factors.

To determine the causes of this spatial and temporal variability in subsurface drainage patterns, GIS data layers were developed. Cluster analysis was used to group the drainage data into meaningful groups to study the factors responsible for such variability. The extrinsic factor of climate was removed through the normalization technique so that the effect of intrinsic factors (soil and landscape attributes) could be analyzed. Cluster analysis offers a means to construct conceptual schemes for organizing information to assist in analysis and to reduce the complexity of a set of data (Bari, 1992). The idea of clustering is to summarize information into interpretable zones and then investigate the causes for spatial occurrence of such zones. Subsurface drainage data were grouped into clusters such that the data within each cluster were similar in some respect but different from those in other clusters.

The number of clusters and their goodness were checked using the cluster evaluation criteria (SAS, 2000). The formation of three clusters showed the highest values of the pseudo−F statistic, R^2, and the cubic clustering criteria. These clusters (1, 2, and 3) were considered as low, medium, and high drainage areas (fig. 1). The spatial occurrence of clusters was not random but seemed to be affected by the topographic attributes and the soil type. The site digital elevation model (DEM) (fig. 2) showed a ridge running along the southeast to northwest corner of the site, which also represented the low and medium drainage areas. The elevation ranged from a low of 25 m (southwest and northeast corners of the site) to a high of 32 m (table 2) in the central area of the site in the northwest to southeast directions. The slope data layer, derived from the DEM, ranged from a low of 1° to a high of 5° (fig. 2). The areas showing higher

Table 2. Descriptive statistics of soil attributes and average normalized crop yield.

Statistic	Elevation (m)[a]	Slope (°)	Aspect (°)	Average Normalized Yield
Mean	28.9	1.7	182	−0.04
Median	29.6	1.5	187	−0.06
Standard deviation	1.6	0.7	37	0.5
Skewness	−0.9	0.7	−0.3	0.4
Kurtosis	0.1	0.3	−0.7	−0.03
Minimum	25	0.6	72	−1.0
Maximum	32	5.0	360	0.9
Interquartile range	2.1	1.0	61	0.6
CV (%)	5.4	39.2	20.4	−−

[a] ≤27 m = low, 27 to 29 m = medium, and ≥31 m = high.

elevations were under a flat slope representing the drainage divide. The patterns of the slope data layer showed resemblance to the patterns of the elevation data layer. The aspect data layer, derived from the DEM, followed the ridge line separating the low and high aspect values in the center of the area and showed a clear different direction of slope on both sides of the drainage divide (fig. 2). The areas on the northeast side of the ridge mostly faced aspect values of 72° from true north. Similarly areas on the southwest of the ridge faced aspect values in the range of 287° to 360° from true north (fig. 2). These topographic data layers gave a very clear indication of the flow paths followed by water based on slope and aspect attributes. This indicates that study of topographic attributes is important to understand and interpret the occurrence of subsurface drainage patterns under artificially drained soils.

GIS Data Layers

The spatial data layers, generated for each year using normalized drainage data, were divided into five drainage classes to study the details of drainage spatial variability patterns over the study period (fig. 3). The areas showing higher drainage (greater than 1 standard deviation) were consistent from 1993 to 1998 and clearly indicated that the northeast corner and central west areas of the site were resulting in higher subsurface drain flows during the entire

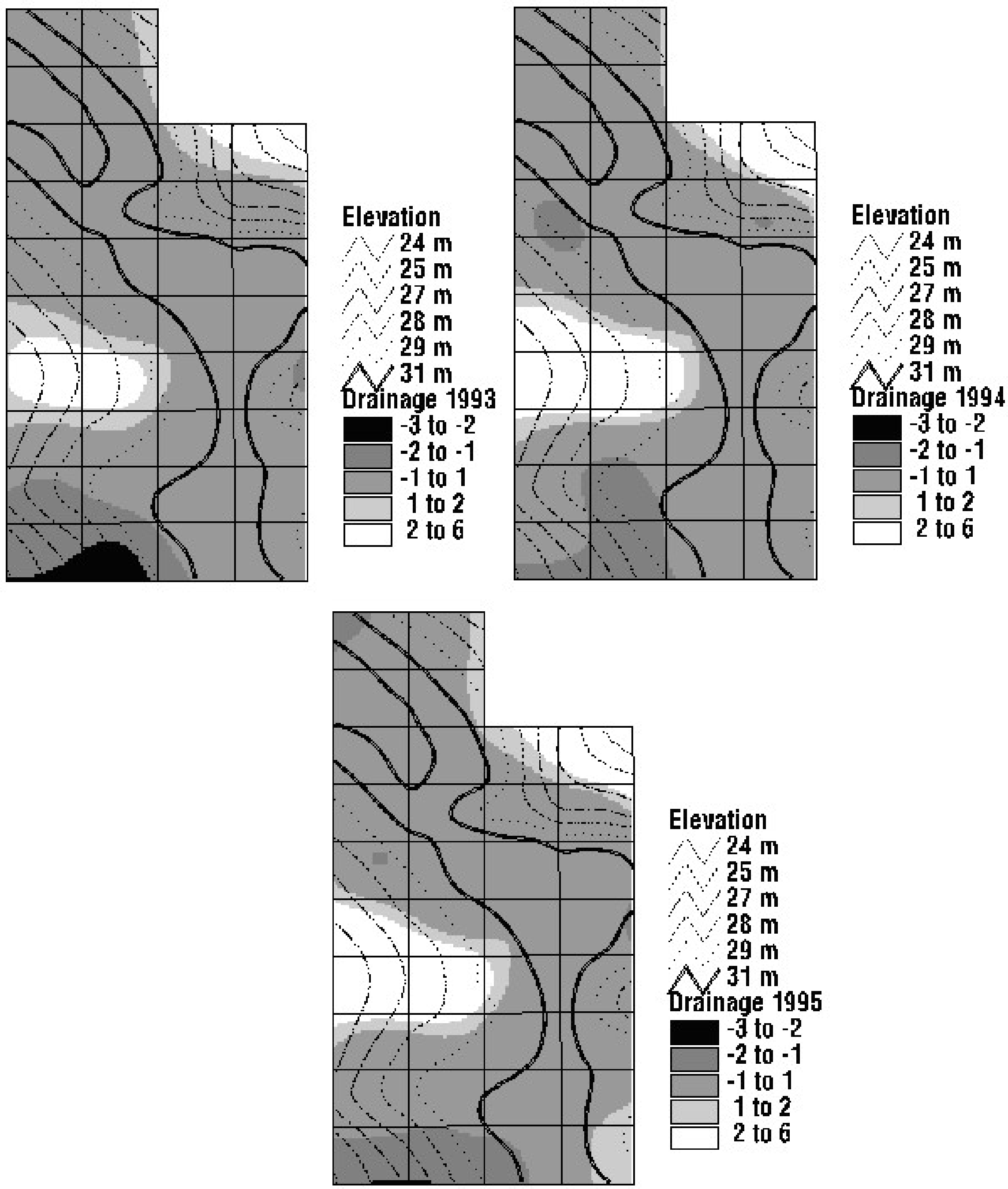

Figure 3. Map overlay of topography and normalized drainage data layers for 1993, 1994, and 1995.

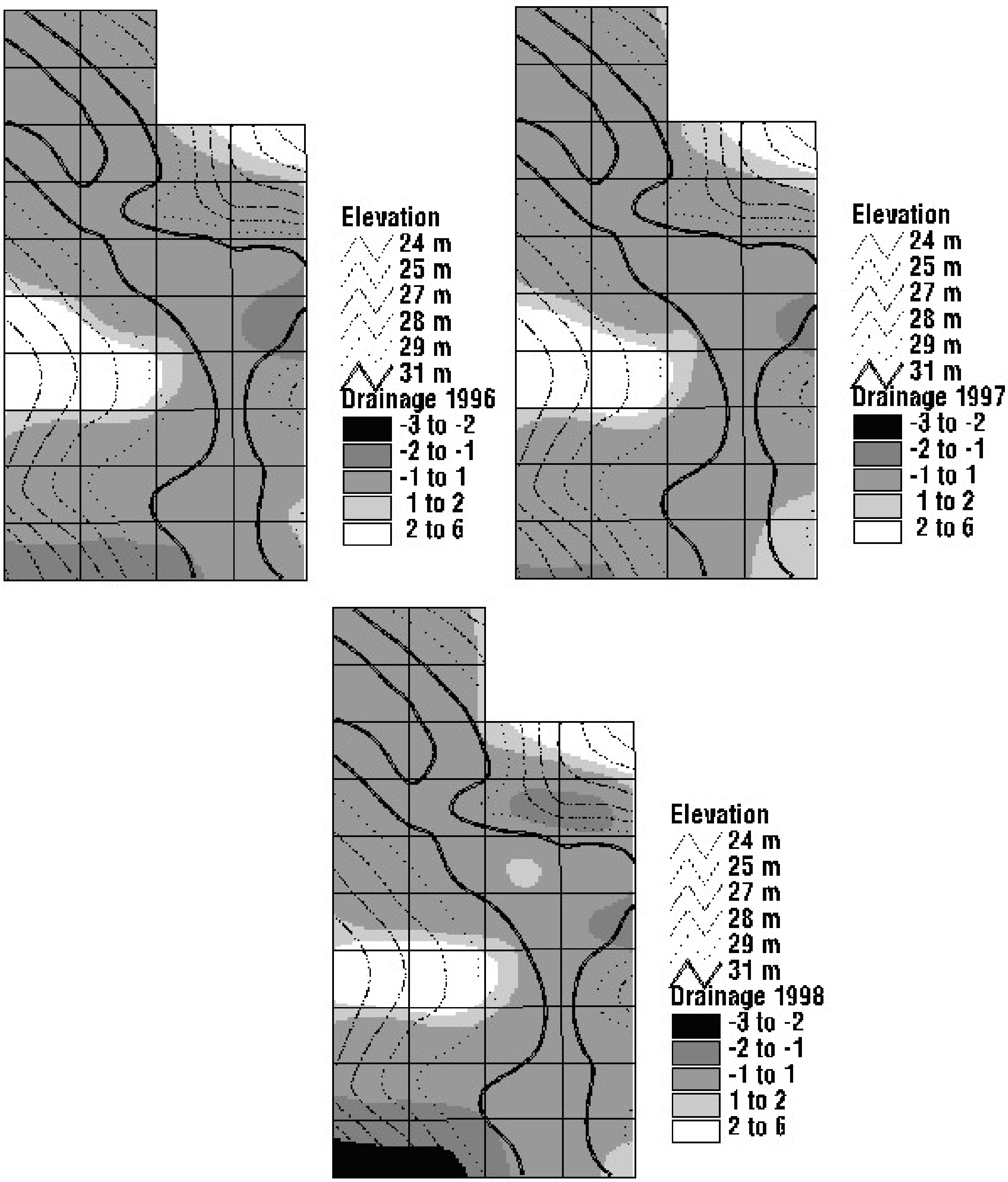

Figure 4. Map overlay of topography and normalized drainage data layers for 1996, 1997, and 1998.

study period (fig. 4). Overlay of plot layout and DEM showed that these high−drainage areas were under low elevation and in a Floyd soil area. These GIS data layers gave a clear understanding of the spatial and temporal variability of the drainage patterns over the 6−year period. In addition, all the 6−year drainage data layers were integrated to determine the common areas under higher drainage using the map query option in ArcView (fig. 5). This integrated theme of high−drainage area was overlaid by soil type and elevation map to analyze its spatial relationships. This analysis also showed that high−drainage areas were located at low elevation levels in the vicinity of Floyd soils (fig. 5).

Cluster Analysis

Seventeen of the 36 field plots were assigned to cluster 1 with a mean value of −0.54 and standard deviation (SD) of 0.48 (table 3). This represented 47% of the area. Cluster 1 was categorized as areas having low drainage. Fifteen of 36 plots were assigned to cluster 2, considered the medium−drainage areas with a mean value of 0.40 and SD of 0.43. Cluster 2 represented 42% area of the site. Cluster 3 had four plots with a mean drainage value of 3.71 and SD of 1.19. These plots were considered high−drainage areas, which were about 11% of the study area. The drainage plots in cluster 3 (high−drainage areas) were located at the lowest elevation levels and in the vicinity of Floyd soils (fig. 1). The medium−drainage

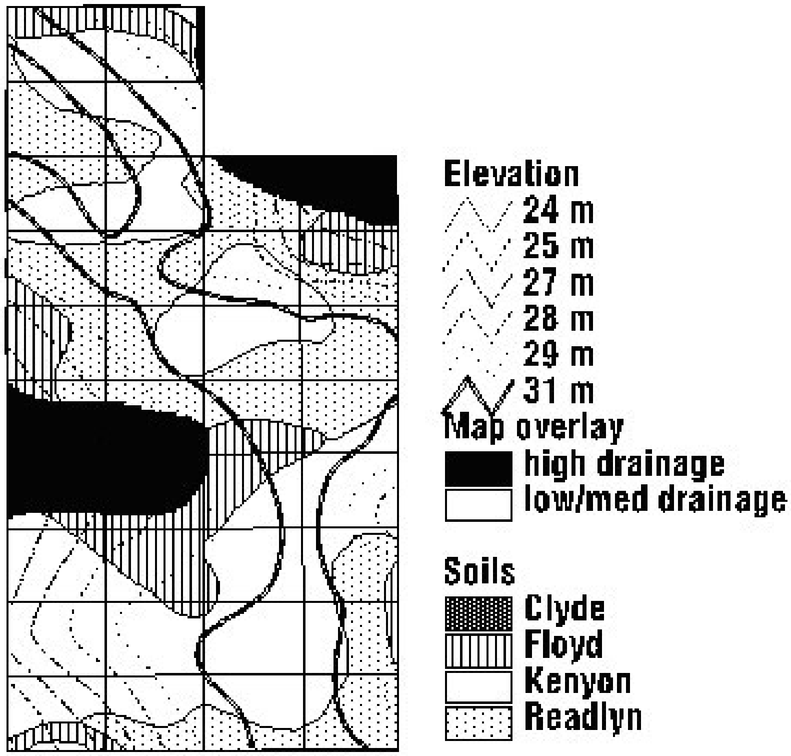

Figure 5. ArcView map overlay output: map overlay of soil type, topography, and the high drainage area, integrated over six years (1993 to 1998).

Table 3. Cluster summary.

Cluster[a]	Mean[b]	Standard Deviation	Weight/ Frequency	Proportion
1	−0.54	0.48	17	0.47
2	0.40	0.43	15	0.42
3	3.71	1.19	4	0.11

[a] 1 = low, 2 = medium, and 3 = high drainage areas.
[b] Normalized subsurface drainage data.

areas (cluster 2) were located around the periphery of high−drainage areas (cluster 3) and on the drainage divide line (fig. 6) where the slope was flat, which might have reduced runoff and increased infiltration. The low−drainage areas in the south of the site were under steep slope varying from 1° to 4° (figs. 2 and 7) and were probably affected by runoff. The spatial representation of cluster analysis clearly showed the topographic effects on the formation of clusters. The northeast and central west sections showed high−drainage areas. This analysis showed the pronounced effect of topography on the drainage patterns of this field. The contribution of various other factors in the formation of clusters was studied using discriminant analysis.

Discriminant Analysis

Discriminant analysis was used to verify the ability of the factors, selected during the stepwise discriminant procedure, to predict membership of each cluster (Jaynes et al., 2003; Li et al., 1992) based on elevation, slope, aspect, soil type, and average normalized yield data. In addition, a pairwise distance matrix showed that the distance between cluster 3 and the rest was quite large in comparison to the distance between clusters 1 and 2 (table 4). The stepwise discriminant analysis, based on clusters, selected elevation, slope (Kravchenko et al., 2002), and the average normalized yield and showed that these variables contributed significantly ($P < 0.15$) to the formation of clusters. These selected variables were used during the discriminant analysis, and it was found that better results were obtained for all three clusters rather

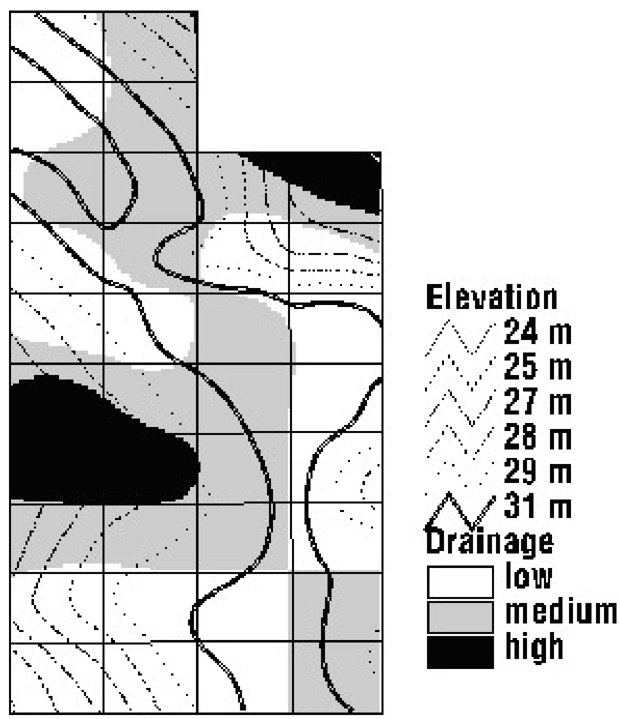

Figure 6. Cluster analysis output: map overlay of topography and drainage clusters.

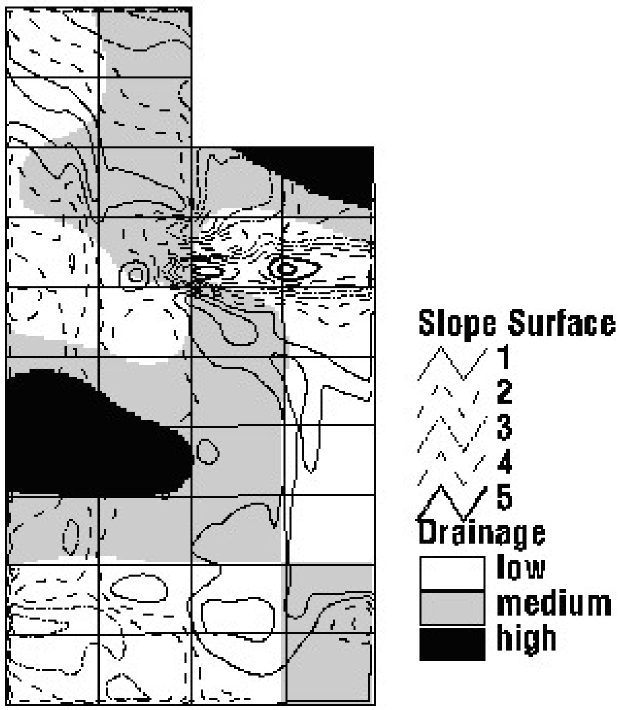

Figure 7. Map overlay of slope surface (deg.) and drainage clusters.

than using all the variables. When the soil type variable, however, was incorporated into the model, cluster 3 was predicted accurately, and all of its four members were assigned to it with zero error. This showed that soil type had a pronounced effect only in the high−drainage areas and was not a significant factor for the low− and medium−drainage clusters. In addition, the entry statistics (table 5) showed that elevation and soil type were the most significant variables qualifying for their entry into the stepwise discriminant model. However, at the end of the discriminant analysis, elevation, slope, and normalized yield formed the stepwise discriminant model, and no other variable qualified to enter the model ($P < 0.15$) (table 6).

The discriminant analysis showed that seven of the 17 members of cluster 1 were accurately assigned to cluster 1, eight members were wrongly assigned to cluster 2, and two were assigned to cluster 3, giving a maximum error rate of 58% (table 7). This was the most poorly defined cluster based on the selected variables. The squared distance between clusters 1 and 2 was also found to be less in comparison to the distance between clusters 1 and 3 (table 4). Ten of 15 cluster 2 members were assigned to cluster 2, and five were assigned to cluster 1. The distance between clusters 2 and 3 was maximum, and therefore no member of cluster 2 was assigned to cluster 3. The error rate of defining cluster 2 was found to be 33%. The most accurately assigned cluster was 3. This cluster had four members. Three of the four were assigned accurately to cluster 3, and one was assigned to cluster 2. The error rate ranged from 25% for cluster 3 to 58% for cluster 1. Similar results have been reported by Jaynes et al. (2003), Kravchenko et al. (2002), and Li et al. (1992) regarding the ability of discriminant functions to predict cluster members.

The variables selected during the stepwise discriminant analysis of elevation (fig. 6), slope (fig. 7), and normalized

Table 4. Pairwise generalized squared distance between clusters.

Cluster	1	2	3
1	0	0.15	7.8
2	0.15	0	9.7
3	7.8	9.7	0

Table 5. Stepwise discriminant model variables entry statistics.

Variable	R^2	F	Pr > F
Elevation	0.34	8.4	<0.01
Soil	0.28	6.6	<0.01
Slope	0.03	0.5	0.60
Aspect	0.03	0.6	0.52
Normalized yield	0.11	2.2	0.12

Table 6. Stepwise discriminant procedure for model formulation based on clusters.

Model Attribute	Attributes Entered	Partial R^2	F	Pr > F
Step 1				
1	Elevation	0.34	8.4	<0.01
Step 2				
2	Elevation	0.48	15.1	<0.01
	Slope	0.25	5.3	0.01
Step 3				
3	Elevation	0.49	15.3	<0.01
	Slope	0.25	5.2	0.01
	Normalized yield	0.14	2.5	0.09

No other variable qualified to enter

Table 7. Confusion matrix.

From Cluster	To Cluster (number of samples/percentage)			
	1	2	3	Total
1	7 (41%)	8 (47%)	2 (12%)	17
2	5 (33%)	10 (67%)	0	15
3	0	1 (25%)	3 (75%)	4
Total	12 (33%)	19 (53%)	5 (14%)	36
Error rate	58%	33%	25%	39%

yield (fig. 8) were used to generate GIS data layers and were overlaid on the cluster analysis output to see their spatial relationship. The patterns of elevation, slope, and yield were closely related to the drainage cluster formations. In this study, it was found that higher drainage areas need better nutrient management plans because of their capacity to discharge higher drainage flow volumes and potential amounts of agricultural chemicals. The GIS analysis showed that the high−drainage areas in both the northeast and central west sections of the site were formed at the lower elevation levels and in the vicinity of Floyd soils. In the stepwise discriminant analysis, it was determined that incorporation of the soil variable in the discriminating function defined cluster 3 very accurately. The success rate was 100% for cluster 3 when soil was included in the discriminant model, but it affected the prediction accuracy for the other cluster groups. Moreover, the contribution of various soil and landscape factors in discriminating cluster groups could differ for different clusters, as was the case for cluster 3. Different discriminating functions developed for every cluster can improve the accuracy of prediction of the member class. The discriminant analysis also showed that more soil and landscape variables (such as

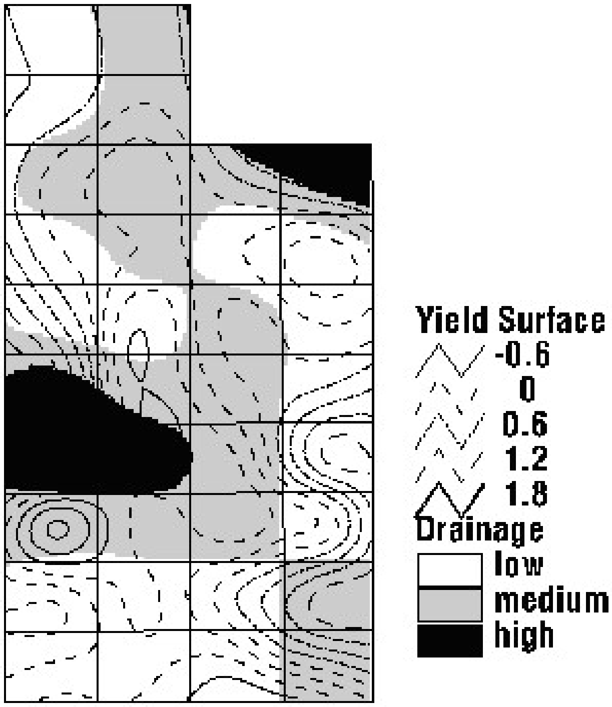

Figure 8. Map overlay of normalized average yield surface and drainage clusters.

organic matter, nutrient profile, soil texture, tilth index parameters, and flow paths) could be considered to predict the membership of each cluster group accurately because subsurface drainage is the integrated outcome of many soil and landscape factors. The analysis conducted in this study showed that soil type was only effective in defining cluster 3 and not the rest of the clusters. GIS analysis, however, showed the integrated effects of elevation level, slope, and normalized yield data on the drainage patterns and determined that interaction between the Floyd soil and low elevation levels contributed significantly to the formation of high−drainage areas.

SUMMARY AND CONCLUSIONS

A technique was developed for the quantitative assessment of subsurface drainage patterns to help in developing management systems to minimize the adverse effects of subsurface drainage water from agricultural fields. Six years of field−measured data on subsurface drain flows were used to develop GIS data layers of drainage patterns and study the relationships among soil and landscape attributes and subsurface drainage data. Patterns of homogeneity were delineated using cluster analysis, and a stepwise discriminant procedure identified elevation, slope, and average normalized yield data as the factors contributing significantly in discriminating these clusters. Discriminant functions developed using the selected variables showed that more soil and landscape variables are required to better predict cluster membership. Map overlay of soil type and GIS data layers of the variables selected during stepwise discriminant analysis concluded that high−drainage areas were located at low elevation levels in the vicinity of Floyd soils of the site. The areas discharging high subsurface drainage flows were identified in the field using this technique and will help the producers/farmers to develop better management practices for these areas to reduce discharge of agricultural chemicals through subsurface drainage water. The combined use of cluster analysis and GIS was found to be useful to group data on subsurface drainage into homogenous zones and study the spatial relationships between these zones and the soil and landscape attributes.

REFERENCES

Al−Abed, S. R., D. T. Lewis, and S. A. Samson. 1989. Use of color splitting and a computer technique to separate soil moisture groups. *SSSA J.* 53(6): 1812−1818.

Al−Sulaimi, J., F. J. Khalaf, and A. Mukhopadhyay. 1997. Geomorphological analysis of paleo drainage systems and their environmental implications in the desert of Kuwait. *Environ. Geol. Berlin* 29(1/2): 94−111.

Ayana, A., and E. Bekele. 1999. Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* L.) Moench) germplasm from Ethiopia and Eritrea. *Genet. Resour. Crop Evol.* 46(3): 273−284.

Ayars, J. E., and D. W. Meek. 1994. Drainage load−flow relationships in arid irrigated areas. *Trans. ASAE* 37(2): 431−437.

Bakhsh, A., D. B. Jaynes, T. S. Colvin, and R. S. Kanwar. 2000a. Spatio−temporal analysis of yield variability for a corn−soybean field in Iowa. *Trans. ASAE* 43(1): 31−38.

Bakhsh, A., T. S. Colvin, D. B. Jaynes, R. S. Kanwar, and U. S. Tim. 2000b. Using soil attributes and GIS for interpretation of spatial variability in yield. *Trans. ASAE* 43(4): 819−828.

Bakhsh, A., R. S. Kanwar, T. B. Bailey, C. A. Cambardella, D. L. Karlen, and T. S. Colvin. 2002. Cropping system effects on NO_3−N loss with subsurface drainage water. *Trans. ASAE* 45(6): 1789−1797.

Bari, M. F. 1992. Cluster analysis as an aid in water resources planning and decision making. *J. Irrig. Eng. Rural Plan*. (22): 29−42.

Brejda, J. J., D. L. Karlen, J. L. Smith, and D. L. Allan. 2000. Identification of regional soil quality factors and indicators: II. Northern Mississippi Loess Hills and Palouse Prairie. *SSSA J.* 64(6): 2125−2135.

Colvin, T. S., D. B. Jaynes, D. L. Karlen, D. A. Laird, and J. R. Ambuel. 1997. Yield variability within a central Iowa field. *Trans. ASAE* 40(4): 883−889.

Corwin, D. L., M. L. K. Carrillo, P. J. Vaughan, J. D. Rhoades, and D. G. Cone. 1999. Evaluation of a GIS−linked model of salt loading to groundwater. *J. Environ. Quality* 28(2): 471−480.

Cressie, N. A. C. 1993. *Statistics for Spatial Data*. New York, N.Y.: John Wiley and Sons.

ESRI. 1996. *ArcView Spatial Analyst.* Redlands, Cal.: Environmental Systems Research Institute.

Feldman, S. B., L. W. Zelazny, and J. C. Baker. 1991. High−elevation forest soils of the southern Appalachians: I. Distribution of parent materials and soil−landscape relationships. *SSSA J.* 55(6): 1629−1637.

Gengerelli, J. A. 1963. A method of determining subgroups in a population and specifying their membership. *J. Psychology* 55: 457−468.

Goolsby, D. A.; W. A. Battaglin, ,B. T. Aulenbach, and R. P. Hooper. 2001. Nitrogen input to the Gulf of Mexico. *J. Environ. Quality* 30(2): 329−336.

Hatfield, J. L., D. B. Jaynes, and J. H. Prueger. 1998. Environmental impacts of agricultural drainage in the Midwest. In *Drainage in the 21st Century: Food Production and the Environment. Proc.*

7th Annual Drainage Symp., 28−35. L. C. Brown, ed. St. Joseph, Mich.: ASAE.

Jardine, N., and R. Sibson. 1971. *Mathematical Taxonomy.* New York, N.Y.: John Wiley and Sons.

Jaynes, D. B. and D. J. Hunsaker. 1989. Spatial and temporal variability of water content and infiltration on a flood irrigated field. *Trans. ASAE* 32(4): 1229−1238.

Jaynes, D. B., J. L. Hatfield, and D. W. Meek. 1999. Water quality in walnut creek watershed: herbicides and nitrate in surface waters. *J. Environ. Quality*28(1): 45−59.

Jaynes, D. B., T. C. Kasper, T. S. Colvin, and D. E. James. 2003. Cluster analysis of spatiotemporal corn yield patterns in an Iowa field. *Agron. J*. 95(3): 574−586.

Jones, J. A. 1989. Environmental influences on soil chemistry in central semiarid Tanzania. *SSSA J.* 53(6): 1748−1758.

Kanwar, R. S., T. S. Colvin, and D. L. Karlen. 1997. Ridge, moldboard, chisel, and no−till effects on subsurface drainage water quality beneath two cropping system. *J. Prod. Agric.* 10(2): 227−234.

Kanwar, R. S., and A. Bakhsh. 2001. Preferential flow and its effects on nitrate and herbicide leaching to groundwater under various tillage systems. In *Preferential Flow Water: Movement and Chemical Transport in the Environment. Proc. 2nd Intl. Symp*., 165−168. D. Bosch, and K. King, eds. St. Joseph, Mich.: ASAE.

Kanwar, R. S., D. Bjorneberg, and D. Baker. 1999. An automated system for monitoring the quality and quantity of subsurface drain flow. *J. Agric Eng. Res*. 73(2): 123−129.

Karlen, D. L., E. C. Berry, T. S. Colvin, and R. S. Kanwar. 1991. Twelve−year tillage and crop rotation effects on yield and soil chemical properties in northeast Iowa. *Commun. Soil Sci. Plant Anal*. 22(19&20): 1985−2003.

Kravchenko, A. N., and D. G. Bullock. 2000. Correlation of corn and soybean grain yield with topography and soil properties. *Agron. J.* 92(1): 75−83

Kravchenko, A. N., G. A. Bollero, R. A. Omonode, and D. G. Bullock. 2002. Quantitative mapping of soil drainage classes using topographical data and soil electrical conductivity. *SSSA J*. 66(1): 235−243.

Li, G. C., R. L. Mahler, and D. O. Everson. 1992. Micronutrients in the Kootenai River Valley of northern Idaho: II. Use of cluster and discriminant analyses to evaluate soil micronutrient status. *Commun. Soil Sci. Plant Anal*. 23(11/12): 1179−1194.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. *Multivariate Analysis.* London, U.K.: Academic Press.

Norton, S. B., S. M. Cormier, M. Smith, and R. C. Jones. 2000. Can biological assessments discriminate among types of stress? A case study from the eastern corn belt plains ecoregion. *Environ. Toxicol. Chem*. 19(4.2): 1113−1119.

Rabalais, N. N., R. E. Turner, D. Justic, Q. Dortch, and W. J. Wiseman. 1999. Characterization of hypoxia: Topic 1 report for the integrated assessment of hypoxia in the northern Gulf of Mexico. Decision Analysis Series No. 18. Silver Spring, Md.: NOAA Coastal Ocean Program.

Rafaelli, S. G., D. R. Montgomery, and H. M. Greenberg. 2001. A comparison of thematic mapping of erosional intensity to GIS−driven process models in an Andean drainage basin. *J. Hydrology* 244(1/2): 33−42.

Randall, G. W. 1998. Implications of dry and wet cycles on nitrate loss to subsurface tile drainage. In *Drainage in the 21st Century: Food Production and the Environment. Proc. 7th Annual Drainage Symposium*, 53−60. L. C. Brown, ed. St. Joseph, Mich.: ASAE.

Randall, G. W. and D. J. Mulla. 2001. Nitrate−N in surface waters as affected by climatic conditions and agricultural practices. *J. Environ. Quality* 30(2): 337−344.

SAS. 2000. SAS OnlineDoc, version 8. Cary, N.C.: SAS Institute, Inc.

Seelig, B. D., J. L. Richardson, and R. E. Knighton. 1991. Comparison of statistical and standard techniques to classify and delineate sodic soils. *SSSA J.* 55(4): 1042−1048.

Singh, P. 1994. Modification of root zone water quality model (RZWQM) to simulate the tillage effects on subsurface drain flows and NO_3−N movement. PhD diss. (diss. abstract 94−24259). Ames, Iowa: Iowa State University.

Voy, K. D. 1995. Soil survey of Floyd County, Iowa. Ames, Iowa: USDA−SCS in cooperation with Iowa State University Cooperative Extension Service.

Wang, X., and Z. Y. Yin. 1998. A comparison of drainage networks derived from digital elevation models at two scales. *J. Hydrology* 210(1/4): 221−241.

Wu, B. M., A. H. C. Van Bruggen, K. V. Subbarao, and G. G. H. Pennings. 2001. Spatial analysis of lettuce downy mildew using geostatistics and geographic information systems. *Phytopathology* 91(2): 134−142.

Young, F. J., and R. D. Hammer. 2000. Defining geographic soil bodies by landscape position, soil taxonomy, and cluster analysis. *SSSA J*. 64(3): 989−998.