IDENTIFICATION OF MANAGEMENT ZONES BASED ON COTTON YIELD AND SOIL PARAMETERS

J. Ping and C. J. Green
Texas Tech University
Lubbock, TX

Abstract

Management zones are the blue-prints for site-specific management practices. However, since crop yields are comprehensively affected by many factors, development of these management zones can be difficult. Furthermore, factors affecting yield can change spatially and may interact with each other, which intensifies the difficulty in developing reasonable management zones. Previous research has shown that delineating management zones is crucial to precision farming. Although several methods have been tested for defining management zones, currently there is no single method in use to meet the practical need for site-specific management. The objective of this paper is to delineate potential management zones from cotton yield and soil parameters by using cluster analysis, discriminant analysis, and other multivariate analysis techniques. Four and three management zones were identified based on the yield groups and MSLCs of soil parameters influencing cotton yields with cluster grouping method and quartile grouping method, respectively. The results also show that the cluster grouping method is better than the quartile methods for delineating yield groups and potential management zones. The application of multivariate analysis can simplify multiple variables and help reveal the relationships between crop yields and soil parameters.

Introduction

An important issue for the successful application of site-specific management is the determination of management zones, which have been defined as "The regions of a farm field that have been differentiated from the rest of the field for the purpose of receiving individual management attention" (Watermerier, 2000). Doerge (2000) defines management zones as, "A portion of a field that expresses a homogeneous combination of yield-limiting factors for which a single rate of a specific crop input is appropriate". Only with scientifically defined and reliable management zones can precision agriculture play its role for increasing production efficiency and reducing potential adverse impacts on the environment.

Currently, spatial variability of crop yield and environmental factors influencing yields are the main basis for defining management zones. Soil surveys, yield mapping, and soil fertility management based on grid soil sampling have been used to make potential management zones (Fridgen et al., 2000). It has been shown that high, medium, and low productivity management zones developed by farmers were effective in developing VRT maps (Fleming et al., 1998). In a study conducted by English et al. (1999), variable potassium input was based on soil nutrient map and potential yield levels. Yang et al. (1999) made variable rate fertilizer application by using the differences between required N and P and the amount supplied by soil based on grid sampling.

Multivariate analyses such as cluster analysis, discriminant analysis, and multivariate analysis of variance have been shown to be helpful in understanding the complex nature of multivariate relationships. Cluster analysis can find the natural grouping of variables by their similarities or differences (Johnson and Wichern, 1998). Discriminant analysis can be used to confirm the classification (Webster and Burrough, 1974). For example, Brown (1991) compared cotton cultivars from different regions more intuitively by means of cluster analysis. In a fruit tree study, discriminant analysis has been shown to be a helpful tool to identify the best conditions for the tree growth (Bonifacio et al., 2000). Fraisse et al. (2000) indicated that cluster analysis and principle component analysis are promising tools for delineating within-field zones that are subject to similar yield limiting factors. In general, cluster and discriminant analysis can be used in pair to find the group for a given situation.

There is an increasing interest in developing groups of crop yields and soil parameters for making reasonable recommendation in crop management systems. The objective of this paper is to delineate potential management zones from cotton yield and soil parameters by using cluster analysis, discriminant analysis, and other multivariate analysis techniques.

Materials and Methods

This research was conducted on a 120 acre production cotton field near Lubbock, Texas during 2000. Three soil series are present and their properties were described previously (Ping and Green, 1999). PayMaster-2200 was planted on May 15, 2000 at approximately 62000 plants/acre. Row spacing was 32 inches. This field was uniformly fertilized with 60 lb of nitrogen and 40 lb of phosphorus per acre. Total growing season rainfall was 11.3 inches, of which 9.2 inches was received between May 25 and June 30. There was no rain in July and August. The cotton received three inches of irrigation water during growing season.

Soil sampling and cotton harvesting procedures have been described previously (Ping and Green, 2000). This season data were collected at 62 sampling points. These points were developed from two sampling schemes: 39 points were collected from the center of a 2.5 acre grid (regular points), and another 23 points were collected from the center of four neighboring regular points (cross points). The soil data used for this paper was the average of the first two layers (12 inches). The data were processed with correlation coefficient analysis, cluster analysis (CA), multivariate analysis of variance (MANOVA), principle component analysis (PCA), and discriminant analysis (DA) through the appropriate SAS procedures. Spatial variability maps were processed with inverse distance weight (IDW) by using Arc Info® and ArcView®. The latitude and longitude degrees in Geographic Reference System from GPS were converted to coordinate system in the unit of meter with the projection of the Universal Transverse Mercator (UTM).

Result and Discussion

Cotton Lint Yields in 2000

Lint yield averaged 595 lb/a and had a CV of 0.22 (Figure 1). Average lint yield decreased by 201 lb/a (25%) compared to 1999 (Ping and Green, 2000). The CV was higher than those in 1998 and 1999 in the same field. The decrease in yield and increase in CV could be related to unusually dry weather from July though September and limited irrigation capacity. Extremely dry weather can increase the yield variation since it intensifies water availability among different soil types and topography (Ping and Green, 2000).

Yield Spatial Variability in 2000

To make comparison with the yield patterns in the previous two years, lint yield was grouped into quartiles. The lowest yield quartiles were located at north and southeast regions (Figure 2). The highest yield regions were located at the west side of the field in 2000. This yield distribution pattern was similar to the one in 1998, which was another dry year (Ping and Green, 1999). However, higher yield regions in 1998 were located in the northwest portions of the field. In 1999, northwest region was one of the lowest regions, which was probably related to the low nitrogen availability caused by leaching or denitrification (Ping and Green, 2000). The input of nitrogen in 2000 may not have been enough to increase the nitrogen to the proper levels, especially where soil nitrate was in extremely low range according to the criteria for cotton production in this region. Nitrogen

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 1:571-574 (2001)
National Cotton Council, Memphis TN

leaching and denitrification may have occurred in the northwest region in June since there were about 9 inches of rainfall around June and this region was a relatively low area in this field. The southeast portion of the field was one of the lowest yield regions in all three years. This area is characterized by high nitrogen, calcium, pH and low phosphorus, and potassium saturation (Ping and Green, 1999; Ping and Green, 2000).

The relationship between cotton lint yields and soil nutrient levels shows that low nitrogen resulted in low yield; whereas relatively high nitrogen with low phosphorus also gave low yield. This was true in the previous two years in this field (Ping and Green, 1999; Ping and Green, 2000). Among three years, lint yields were correlated to soil pH values. Higher sand contents lead to higher lint yields in 1998 and in 2000. However, there was no significant direct relationship between potassium saturation and cotton lint yield in 2000. Due to the complex effects of soil parameters on cotton yields, single correlation analysis may not fully explain the relationship between yield and soil parameters.

Yield Groups Classified by Cluster Analysis

Quartile yield groups were used to further evaluate the effects of soil properties on yields. This method is, however, arbitrary because groups are divided by their ranks. As a result, there tends to be equal numbers of data points in each quartile group. In reality, crop yields tend to cluster together under similar soil and other environmental conditions. Cluster analysis is one of the statistical ways to group variables based on their agronomic similarities.

Based on the dendrogram and scree plot of lint yields over the 62 points from cluster analysis, four groups were selected. Since points within each group are similar to each other, they tend to be similar in soil properties and may be amenable to common management practices (i.e. they may be in the same management zone). Unlike the quartile method of classification, cluster groups would not necessarily contain the same number of data points (Table 1).

The differences in yield between two kinds of grouping methods are listed in Table 1. It seems that grouping with the cluster method can reduce the range and standard deviation in the high and low yield regions compared to those obtained with the quartile method. Because there were less yield points near median range than would be expected for more normally distributed data (Figure 1), there were smaller variations in the regions right above and below the median as defined by the quartile grouping method. It can be predicted that the cluster method should give better performance as group number increases or yields distribute more normally.

Relationships Between Yield Groups and Soil Parameters

With either grouping technique, cotton yields were significantly different among the resulting four groups. ANOVA and further multiple comparisons were conducted with selected soil parameters (Table 2). These variables were evaluated here because they were correlated to cotton yield in 2000 and the previous two years (Ping and Green, 2000).

However, there are no significant differences in these main soil parameters by grouped yields with cluster and quartile methods. The soil parameters listed in Table 2 tend to be significantly different between Group 1, 2 and Group 3, 4. They are similar within Group 1 and 2, and similar within Group 3 and 4. Except soil pH, the soils would be divided into two categories based on these variables in either grouping method. The yield pattern can tell that two categories of soil management zones may not explain yield differences properly. Ideally, one index that can match the yield pattern would be desirable.

Management Zones Based on Soil Parameters from Three Methods

Cotton yields are comprehensively affected by many factors, which could interfere with the interpretations of the effects of these factors on yields. There is a need to find a simpler variable that can represent the multiple variables and visualize the relationship between yield and related variables more intuitively.

Statistically, the first principle component is the linear combination with maximum variance. There exists one linear combination from multivariate analysis that can separate a given class variable and simplify the multiple variables greatly (Johnson and Wichern, 1998). In this study, one principle component and two linear combinations to separate clustered yield groups and quartile yield groups were tested by means of ANOVA and multiple comparison in one dimension. The results are listed in Table 3. There are no significant differences in the first principle components between the two grouping methods. However, the linear combination from the cluster method can group soil parameters in four significantly different categories, which can be used to determine different management practices.

Based on the values of linear combination for each point, three soil management zone maps were created. Four management zones were developed based on the soil parameters grouped by clustered yield groups since there are significant differences in this linear combination (Figure 3). Three management zones are from the quartile grouped yields since only three groups of soil parameters are significantly different (Figure 4). It seems that these two maps are quite similar except that they define different potential management zones in the lowest yield range. Statistically, two management zones should be derived from the first principle component. For making better understanding of overall variations of soil parameters, four management zones were developed. It was obvious that management zones based on the first principle component are quite different from those developed from the other two methods (Figure 5). These management zones relate to the changes in soil series, soil texture and soil calcium contents.

The number of potential management zones from our data set were similar with the zone number developed by Fridgen et al. (2000) from soil EC, elevation, and slope, in which they concluded that no advantage of dividing fields into more than four or five management zones. Increasing the number of management zones within a field may more effectively manage uniformity of crop yields. However, the goal of precision farming management is not to manage uniformity, but to maximize net economic return and increase sustainability.

The potential soil management zones from these linear combinations can be thought of as the overall performances of soil factors related to cotton yields. They can serve as the basic management zones for further variable application experiments. For example, phosphorus effects can be checked among those zones and the responses from variable rates of phosphorus can be combined to develop a variable input of phosphorus application. The zones developed from this year should be compared with other years for making more practical management zones.

Evaluation of Two Methods of Grouping Cotton Yield and Soil Parameters Zones

To compare the effects of cluster and quartile grouping methods, discriminant analysis was conducted. This method can conform if predetermined variables can fall in the correct category based on the values from tested variables.

The results from discriminant analysis are usually expressed as error accounts from predefined categories with cross validation. Higher error rate indicates an improper classification. In this study, two methods were tested with same data sets. Results show that the clustered grouping method has less error than the quartile grouping method (Table 4). This indicated that cluster grouping method is better than quartile grouping method for classifying yield and soil parameter to make potential management zones. This result is consistent with the linear combination that maximally separates the soil parameter by clustered groups.

Summary

With our data, four yield groups were identified with both cluster and quartile grouping methods. Four and three management zones were identified based on the yield groups and MSLCs of soil parameters influencing cotton yields with cluster grouping method and quartile grouping method, respectively. However, cluster grouping method has less errors for yield classifications than quartile grouping method. The application of multivariate analysis can be used to further explore the relationships between cotton yields and environmental factors.

Acknowledgement

Funding for this project was provided by Cotton Incorporated and by the International Cotton Research Center. The GPS receiver used in this project was made available by Oministar In, Houston, TX.

References

Bonifacio, E., S. Santonoi, and E. Zanini. 2000. Soil properties required by some southern Africa fruit trees as assessed by discriminant analysis. Arid Soil Research and Rehabilitation. 14:253-263.

Brown, J. S. 1991. Principle component and cluster analysis of cotton cultivar variability across the U.S. cotton belt. Crop Sci. 31:915-922.

Doerge, T. 2000. Defining management zones for precision farming. http://www.pioneer.com/usa/technology/precision.htm.

English, B. C., R. K. Roberts, D. Sleigh, and S. B. Mahajanashetti. 1999. Spatial break-even analysis of variable rate potassium application technology on cotton: an application on a soybean-cotton field in Western Tennessee. Proceedings – Beltwide Cotton Production Research Conferences. 1:31-32.

Fleming, K. L., Westfall, D. G., D. W. Wines, L. E. Rothe, J. E. Cipra., and D. F. Heermann. 1998. Evaluating farmer developed management zone maps for precision agriculture. *In* Robert, Rust, and Larson (eds). Precision agriculture. Proceedings of the 4th international conference. Minneapolis, Minnesota. p. 335 – 343.

Fraisse, C. W., K. A. Sudduth, N. R. Kitchen, and J. J. Friden. 2000. Use of unsupervised cluster algorithms for delineating within-field management zones. http://www.fse.missoui.edu/mpac/pubs/zones.htm.

Fridgen, J. J., C. W. Fraisse, N. R. Kitchen, and K. A. Sudduth. 2000. Delineation and analysis of site-specific management zones. http://www.fse.missoui.edu/mpac/pubs/zones.htm.

Johnson, R. A. and D. W. Wichern. 1998. Applied multivariate statistical analysis. Prentice Hall., Upper Saddle River, NJ. 07458.

Ping, J. and C. J. Green. 1999. Spatial analysis of agronimic properties in two production cotton fields in West Texas. Proc. Beltwide Cotton Conference 2:1286-1290.

Ping, J. and C. J. Green. 2000. Factors affecting yield variability in irrigated cotton. Proc. Beltwide Cotton Conference. 2:1404-1407.

Watermerier, N. 2000. Managing inputs through crop management zones. http://precision.ag.ohio-state.edu/~precisfm/library/.

Webster, R. and P. A. Burrough. 1974. Multiple discriminant analysis in soil survey. J. of Soil Sci. 25:120-134.

Yang, C., G. L. Anderson, J. H. King, and E. K. Chandler. 1998. Comparison of uniform and variable rate fertilization strategies using grid soil sampling, variable rate technology, and yield monitoring. *In* Robert, Rust, and Larson (eds). Precision agriculture. Proceedings of the 4th international conference. Minneapolis, Minnesota. p. 675 – 686.

Table 1. Differences in yield groups based on cluster and quartile analysis

Statistics	Method	Group 1	Group 2	Group 3	Group 4
Size	Cluster	13	22	22	5
	Quartile	16	15	16	15
Mean	Cluster	419.42	549.57	684.31	854.58
	Quartile	437.69	547.92	631.10	770.07
Std. Dev.	Cluster	45.75	22.12	45.95	55.58
	Quartile	56.76	15.02	35.67	70.87
Min.	Cluster	334.36	510.79	618.28	795.15
	Quartile	334.36	522.03	575.99	688.55
Max.	Cluster	477.31	585.9	752.64	921.59
	Quartile	521.81	574.67	678.85	921.59
Range	Cluster	142.95	75.11	134.36	126.44
	Quartile	187.45	52.64	102.86	233.04

Table 2. Differences in selected soil parameters based on cluster and quartile analysis

Group	Method	pH	Ca (ppm)	Sand (%)	Clay (%)	D to R (cm) †
1	Cluster	8.13a‡	1987a	67.0a	19.0a	62.5a
	Quartile	8.11a	2001a	74.0a	20.6a	63.7a
2	Cluster	8.00ab	1895a	67.6a	20.2a	75.2a
	Quartile	8.03ab	1926a	70.5ab	18.8ab	72.1ab
3	Cluster	7.84bc	1482b	72.6b	17.7b	85.8b
	Quartile	7.92b	1535b	68.2b	18.8ab	77.0ab
4	Cluster	7.70c	1537ab	72.7a	18.4a	78.0a
	Quartile	7.73c	1495b	65.8b	17.4b	93.1b

†= Depth to HCl reaction layer; ‡ Same letters among different yield groups for a given soil parameter are not significant different ($P \leq 0.05$).

Table 3. Comparisons on three linear combination for determining soil management zones

Source	Method	Group 1	Group 2	Group 3	Group 4
MSLC†	Cluster	0.16a	0.04b	-0.08c	-0.21d
	Quartile	0.14a	0.06a	-0.03b	-0.18c
Prin1‡	Cluster	0.59a	0.24ab	-0.47b	-0.54b
	Quartile	0.45a	0.43a	-0.30b	-0.60b

† Most significant linear combination; ‡ Only first principle component was listed here

Table 4. Errors from two grouping methods for cotton yield

Method	Group 1	Group 2	Group 3	Group 4	Total
Quartile	0.12	0.13	0.13	0.20	0.15
Cluster	0.15	0.00	0.05	0.00	0.05

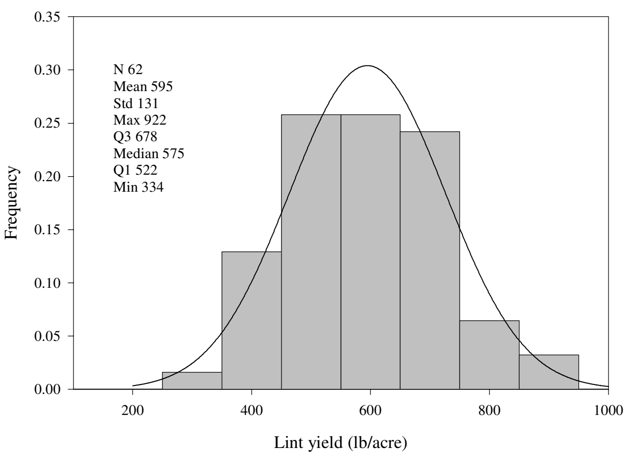

Figure 1. Lint yield frequency distribution and its theoretical normal distribution under same mean and standard deviation

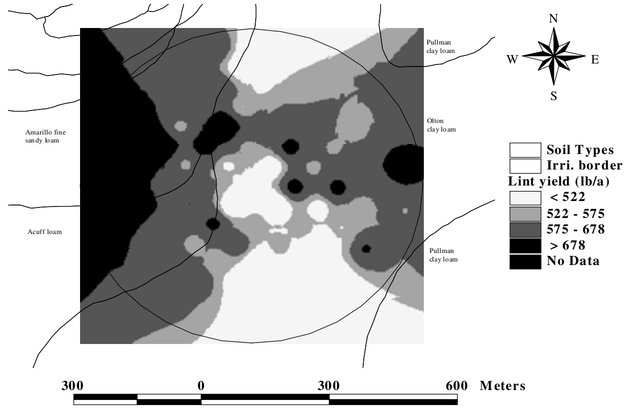

Figure 2. Yield spatial variability based on quartile group

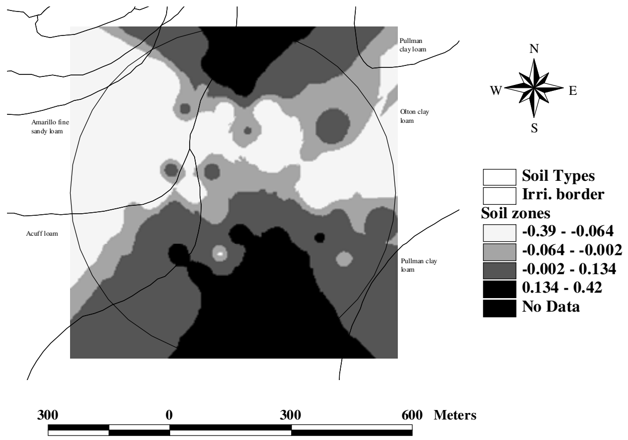

Figure 3. Potential management zones based on cluster yield groups and linear combination of soil parameters maximizing these groups

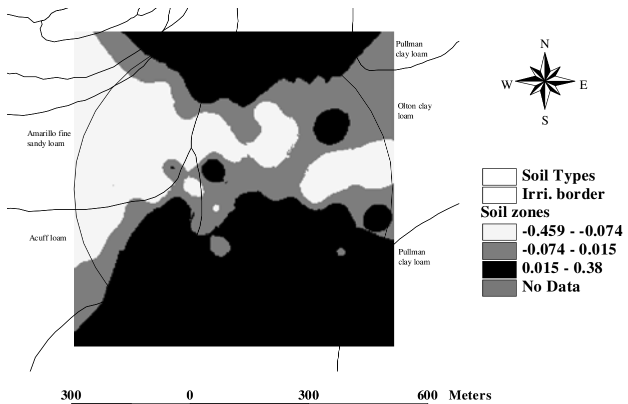

Figure 4. Potential management zones based on quartile yield groups and linear combination of soil parameters maximizing these groups

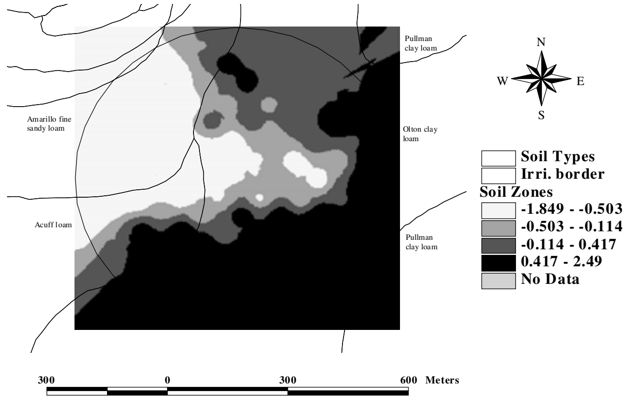

Figure 5. Potential management zones based on first principle component of soil parameters