Agriculture Field Characterization using GIS software and Scanned Color Infrared Aerial Photographs

P. Basnyat[1], B.G. McConkey[1], G. Noble[2] and L.B. Meinert

[1]Semiarid Prairie Agricultural Research Centre, 1 Airport Road, Swift Current, Saskatchewan S9H 3X2
[2]Rural Service Centre, Saskatchewan Agriculture & Food, Box 250, Assiniboia, Saskatchewan S0H 0B0

Abstract

This paper addresses the potential of a color infrared aerial photograph to provide spatially distributed information for site specific management. In this process digitized color infrared aerial photographs were used to extract vegetation index information. In addition, important crop and soil information were also collected using a grid sampling technique. Crop and soil information contributing most to explain variability were determined and used in further analysis. Grain yield data obtained using combine sampling were noted along with the coordinate information of the sample points. Locational information were collected using GPS. Kriged surface were generated using soil and crop point sample information. Point information were extracted from each kriged surface using centroid of uniformly spaced grid (15 m cell). Fuzzy k-means with extragrades algorithms were used to delineate potential within-field management units based on soil and crop information and vegetation index separately. Then "goodness" of potential management zones generated using within zone variability of grain yield. Ideal number of zones were determined using the decrease in total within-zone variance. Finally, management zones determined using crop and soil information and vegetation index information were compared for similarity. The methodology is fast, can be easily automated in commercially available GIS software and has considerable advantages when comparing to other methods for delineating within-field management zones.

Introduction

Site-Specific agricultural management involve collection and control of agronomic information to supply actual needs to parts of fields rather than average needs to whole fields. Early attempts in variable rate application of crop inputs were based on grower intuition, soil survey maps, soil test data from sparsely spaced grid samples, and direct observation of soil conditions while traversing variable fields were of limited success. Continued research has revealed the importance of considering additional site characteristics that exert a major influence on crop yield. The additional information needed for better management can come from extensive soil sampling using a dense grid sampling procedure which are costly and time consuming. Determination of sub-field areas is difficult due to the complex combination of potential yield-limiting factors. Additionally, improperly defining management zones may be no better than uniform management of the field.

Large-scale aerial photographs can assist in field-scale soil mapping (Harrision et al. 1987; Lillesand and Kiefer 1994). Although past use of aerial photographs in crop and soil studies usually involved manual interpretation, computer digitization and spatial registration now allow statistical approaches that can quantify the variability captured with photographs (Lillesand and Kiefer 1994). Yield patterns can be differentiated using aerial photograph of bare and cropped surfaces (Long et al. 1989). Photographs taken at physiological maturity of the crop canopy within uniformly managed fields were useful in pattern differentiation in the crop canopy (Long et al. 1989; Milfred and Kiefer 1976; Lillesand and Kiefer 1994). Photographs of a developing crop can be used to discern management effects, but soil landscape effects on the crop are best discerned at maturity. Historical aerial photographs, if available can yield useful information regarding past practices and cultural operations and past soil and yield information can help in identifying stability of spatial pattern temporally.

Crop and soil mapping using aerial photographs is possible based on two facts of aerial photography. First, growing vegetation absorbs red light (0.6 to 0.7 micron wave lengths) and reflects near infrared light (0.7- to 1.3 micron wave lengths). Color-Infrared Photographs Color films, either standard or infrared, have three layers (emulsions) which, when processed, produce yellow, magenta, and cyan dyes respectively. For standard color film, each layer is sensitive to a finite part of the visible spectrum (red, green, and blue) such that the composite image looks much like the colors our eyes would perceive. For color-infrared film, the emulsions are made sensitive to different bands of the spectrum. In this film the top layer is sensitive to near-infrared light, enabling the film to image this type of light as red in the resulting pictures. Similarly, red will show as green; green as blue; and blue as black. Because infrared film does not produce an image that appears natural, it is often called "false-color"film.

Plant stress and maturation are accompanied by greater reflectance of red light and more absorption of infrared light. A loss of growth and vigor in the crop can be detected by increased red reflectance and lowered infrared reflectance. Monitoring drought stress by satellite imagery has been based on this principle (Lillesand and Kiefer 1994). Similarly, differential crop maturation within fields might also be captured by properly timed color-infrared photographs.

Agricultural remote sensing involving crops and soils is extremely complex because of the dynamic nature and inherent complexity of biological materials and soils, yet remote sensing technology offers numerous advantages over traditional methods of extracting soils and crops related information by the virtue of it’s cost effectiveness and easy and timely data collection and analysis over a large area.

Because of the continuous nature of soils data, classification systems that allow any one observation to belong to exactly one class are often inappropriate. Fuzzy or continuos classification procedures were developed to use in situations where class boundaries are not, nor can they be, sharply defined. In contrast to crisp classification, continuous classification procedures allow individuals to have partial class membership (i.e. an individual can belong to more than once class) (Burrough et al., 1992). Continuous classification has been widely used for soil classification and delineation of management units (McBratney and deGruijter, 1992; Odeh et al., 1992; Boydell and McBratney 1999; Lark 1998).

For site-specific management, careful consideration of the likely management operation or agri-chemical input(s) to be employed is needed in order to determine the procedure of how to divide a field into different management zones. Once the management units are determined, each zone should represent a unique combination of potential yield-limiting factors for which we can improve the site-specific management prescription. Relative to management dependent on soil/landscape variation, the question considered here is "how should different management zones be determined?"

Objectives

The primary objective of this research was to evaluate if data obtained from scanned aerial-infrared color photographs can be used to predict and map variation of wheat grain yield within a managed field. The second objective was to compare management units delineated using information extracted from photograph and information gathered through intensive-sampling. An important aspect of this work was to ensure that observed spatial patterns in crop productivity were consistent with soil and topographical variability.

Materials and Methods

The study site was near highway 13, in Shaunavon (NE10-09-18w3; area 24.09 ha). A color infrared aerial photograph (Figure 1) of this area was taken in 21st August 2000.

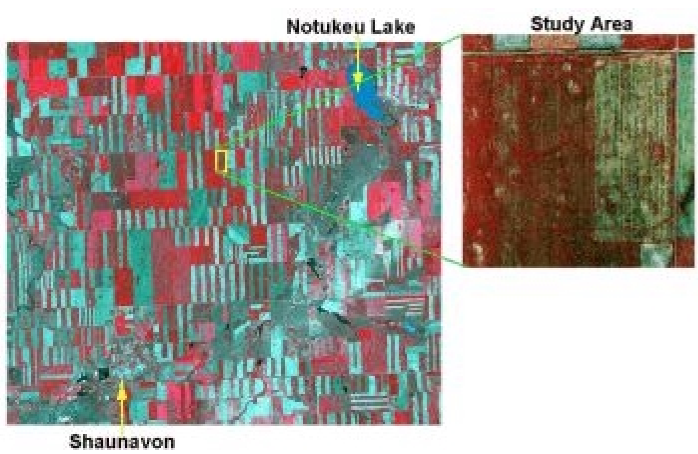

Figure 1: The Study Area

The field was planted with spring wheat under uniform fertilization. Each plot inside the study area (bounded by red polygon, area 0.963 ha) received different amount of fertilizer in previous year (i.e. from 1997 to 1999). Darker red pigmentation on the photograph is the indication of more green vegetation. Topographically the area is more undulated as shown in Figure 2 below. 3D image of both study area was based on the elevation information obtained from the ground control points using geographic positioning systems(GPS).

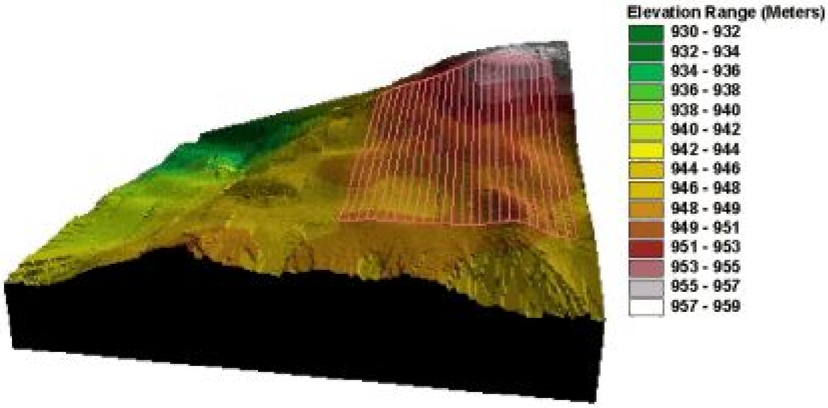

Figure 2. Topography of the study area

Color infrared image of each study area were scanned at 600 dpi using flat-bed scanner to give eight bit (256 levels), three band (RGB) tagged image file format (TIFF) images. The scanner measures relative intensity of light reflected from the photographs after passing filters that allow transmission of red, green and blue wave lengths. Red, green and blue image bands indicate photographic half tones within infra-red , red, and green bands, respectively. Greater half tone value indicate greater reflectance in that band. The resulting image files were imported into IDRISI geographic information system (Clark University, Worcester, MA) where they were band separated and georegistered. Georegistration was carried out using reference points obtained from a global positioning system (GPS). During geo-registration the images were re-sampled so that the number of image cells per crop row width were adjusted to the whole number nearest to the original image.

There are many vegetation indices that could be used to interpret aerial photograph. Almost all of these are based on formulas involving the magnitude of the number representing the red band intensity, denoted R, and that representing the IR band intensity, denoting IR. We have used a Normalized Difference Vegetation Index (NDVI). NDVI is calculated as

$$NDVI = IR – RED / IR + RED$$

The resulting vegetation index for each study area is presented in Figure 3 .

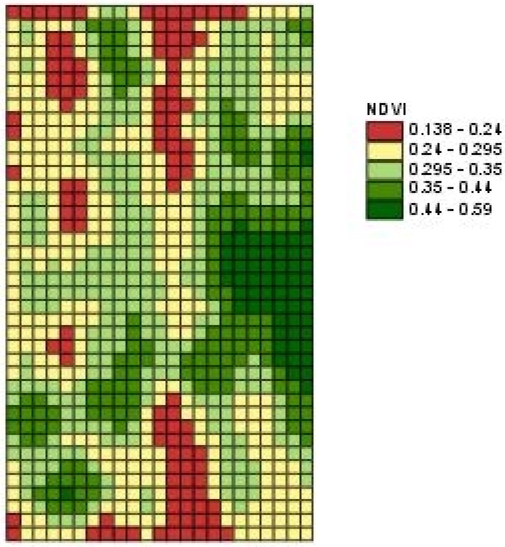

Figure 3. NDVI of the Study Area

Yield Data

Grain yield measurements were obtained using a combine sampling. Mid position of combined area (may be several within each plot based on the landscape position) were noted. A grid surface was generated using kriging from normalized point sample information, normalization was done based on the mean yield of the field. Grid surface reveals the yield pattern of the crop (Figure 4), which shows variation within and between experimental plots. Once the yield information is collected correlation between vegetation index and yield were calculated.

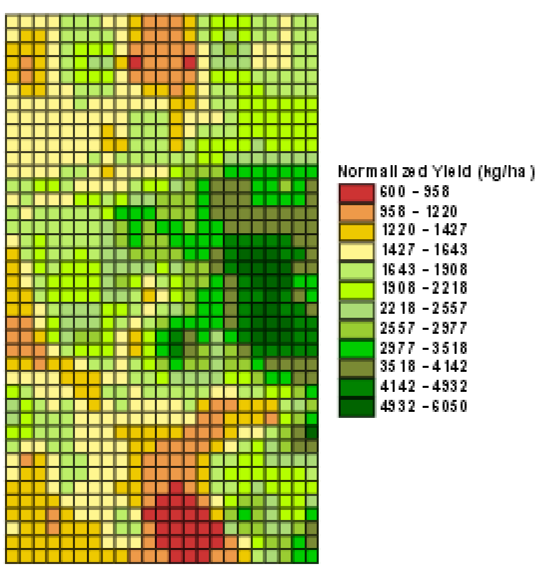

Figure 4. Normalized grain yield

Clustering Process

The main objective of unsupervised classification or clustering is to identify naturally occuring clusters in a data. The clustering algorithm used was the fuzzy k-means with extragrades (deGruijter & McBratney 1988) which is a modification of fuzzy c-means (Bezdek, 1981). This modified algorithm based on the assumption that continuous classes should provide better representations of outliers or atypical individuals than discontinuous classes. This is especially the case with outliers located between clusters in property space. (i.e., intragrades) Fuzzy k-means, for instance, will indeed give intermediate memberships to intragrades. However, outliers outside the main body of data points, referred to as extragrades, are still not suitably represented by fuzzy k-means. deGruijter & McBratney (1988) modified the objective function to account for extragrades. This improvement makes the memberships directly depend upon the distances to the class centroids. The classification was accomplished using the program "FuzME". developed at the Australian Centre for Precision Agriculture (Minasny and McBratney 2000). This software determines membership in each cluster through an iterative process beginning with a random set of cluster means. Each observation is assigned to the closest to these means. New means are recalculated for each cluster based on the distance (Euclidean, Diagonal, or Mahalanobis) from the observation to cluster mean. To ensure cluster stability, the process is repeated until either the specified convergence criterion is met or the maximum number of iterations, stopping criterion, fuzziness exponent, and distance metric. The software generates a data file containing each observation, the cluster to which it belongs, and a vector of membership values to indicate how closely each individual resembles each cluster. To obtain good convergence, a stopping criterion of 0.00001 was used in this study, which means the change in membership with an iteration of means calculation was less than 0.00001. Because, the variables used for classification exhibited significantly different ranges, the distance metric was set as to Mahalanobis. In this mode, the data set is transferred to one in which all attributes have zero mean and unit variance. Correlations between variables are accounted for as well (McBratney and Moore, 1985). The optimum number of zones were decided based on the fuzziness performance index (FPI) and modified partition entropy (MPE). FPI estimates the degree of fuzziness generated by a specified number of classes. MPE estimates the degree of disorganization created by a specified number of classes. The optimum number of classes was established on the basis of minimizing these two measures (Roubens 1982).

Results and Discussion

To better represent the grain yield , it was normalized with the average yield of the entire field, then these normalized point information were used in generating surface using kriging interpolation. Yield variance for the entire filed was calculated as 991624, which was used in the comparison of variance reduction after the field was divided into a number of management zone using clustering process.

Determination of Management Zones

The unsupervised classification procedure was used to delineate management zones beginning with three zones and further dividing into maximum of fifteen zones. As the number of zones increased less pronounced and possibly less interpretable features were included. Figure 4 shows the results obtained when field was classified into thirteen zones using NDVI information. When compared with the surface generated using yield information (Figure 3), we can see a similarity, but the problem may arise in the actual management due to it's salt and pepper nature.

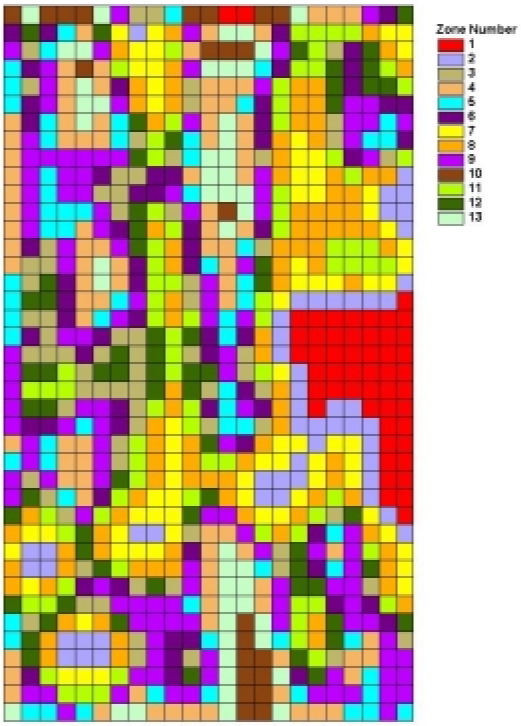

Figure 4. 13 zones delineated using Fuzme2 Information.

Determination of the Optimum Number of Zones

The Fuzme2 software was used to divided the study area into three to sixteen zones using NDVI as the input variables. After calssification, optimum number of management zones were determined based on the FPI and MPE. The FPI and MPE only model the variation in the attributes used to perform the classification which serves as a first step in the determination of the optimum zones. The FPI is a measure of the degree to which different classes share membership and is constrained to values between 0 and 1. As FPI approaches 1, membership sharing increases. As the FPI approaches 0, classes become more distinct with less membership sharing. The MPE is an estimate of the amount of disorganization created by a specified number of classes. Like the FPI, it is also constrained to value between 0 and 1. As MPE approaches 1 , disorganization predominates while values approaching 0 indicate excellent organization. Plotting the values of FPI and MPE against the number of classes and choosing a classification that minimizes both measures was the criteria used in the identification of optimum number of zones. The plots of clustering performance (FPI and MPE) against the number of zones is presented in the Figure 5 and 6 respectively.

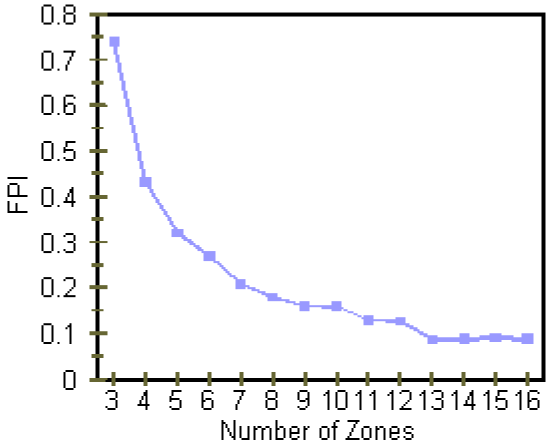

Figure 5. Clustering performance FPI against the number of zones.

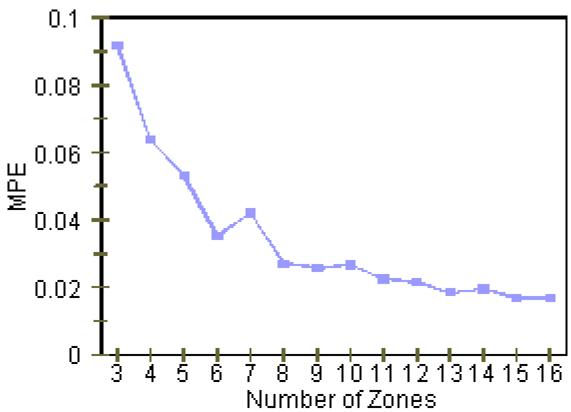

Figure 6. Clustering performance MPE against the number of zones

The result of the plotting of the FPI and MPE against the number of zones indicates poor initial segregation of the data and a large amount of disorganization starting with three zones. The FPI attains a minimum value at 13 zones, but the MPE plot shows the early minimum near five zone, but it eventually decreases after seven zones and reaches minimum at 13 zones. Hence we have checked the performance of both 5 zones and 13 zones. But in the absence of crop information such a s grain yield., the decision criteria for selecting optimum number of zones should be the minimum of both FPI and MPE with a qualifier that if such zones defined are large (depending on the area under consideration) then select the zones which hits early minimum (e.g., in our case looking at the MPE, plot hits initial minimum near zone 5). If the additional variable like grain yield information is available, it will be used to validate the zone identified by the FPI and MPE.

Yield Analysis

Yield information from the study area were used to validate the optimal number of zones identified by the FUZME indices. Yield data were normalized by dividing the measured yield for each grid cell by the mean value for the entire field so the amount of uniformity gained within each zone by dividing the field into additional zones could be determined. If we continued dividing the field into a large number of zones, ultimately there would be as many zones as grid cells and there would be no yield variability within a zone. Fraisse and others (1999) weighted yield variances for each zone based on the area of the zone using:

Where:
S_z^2 = Weighted variance for zone Z;
Y_i = Yield measured for cell I;
$\bar{Y}_z$ = Mean of the measured yield in the zone Z;
n_z = Number of cells in zone Z;
n_T = Total number of cells in the map.

The decrease in total within-zone variance was used to select the most appropriate number of zones. Total within-zone yield variance of a map was defined as the sum of weighted within-zone yield variances for each zone:

$$S_T^2 = S_1^2 + S_2^2 + \ldots\ldots + S_z^2$$

The average yield in 13 zones identified above is shown in Table 1. The superscript number indicates the reclassed number based on the significance of yield variability between each zone.

Table 1. Average grain Yield inside each zone
(For 13 zones identified)

ZONE	YIELD	ZONE	YIELD	ZONE	YIELD
1	4363^1	5	1862^3	9	1833^3
2	3113^2	6	1882^3	10	1162^5
3	1900^3	7	2342^4	11	2129^4
4	1621^5	8	2240^4	12	1862^3
		13	1380^5		

Total within-zone yield variance: 546376 (55% of entire field variance). The percentage decrease in within-field variance obtained by dividing the study area into management zones using fuzzy k)means algorithm is shown in Figure 7. Yield variance for entire field was used as the reference, 100% level. In this case the total variance in yield is within the zone. As the number of zones increases, a portion of the total yield variance is explained by the zone partitioning (between zone variance) causing the within- zone variance to decrease. The early maximum decrease in yield variance was obtained by dividing the field into five. With the increase in the zone number there are further decrease in variance. This plot resembles just like MPE plot.

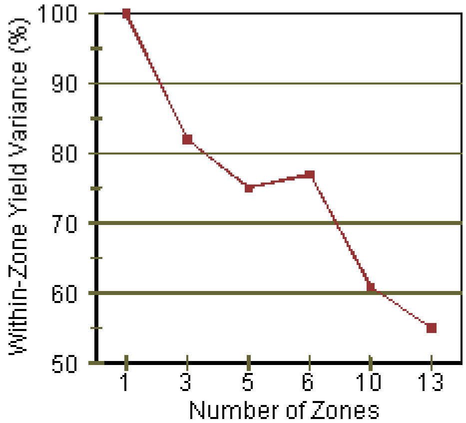

Figure 7. Decrease in within-zone yield variance (%) obtained by dividing field into management zones.

The average yield within each of the five management zones identified by the variance reduction process is given in the Table 2 below. Total within-zone yield variance: 744632 (75% of entire field variance). The management zone identified is presented in Figure 8.

Table 2. Mean Yield inside each zone
(Optimum zone based on variance reduction information)

ZONE	YIELD
1	2357
2	2079
3	1898
4	1962
5	1789

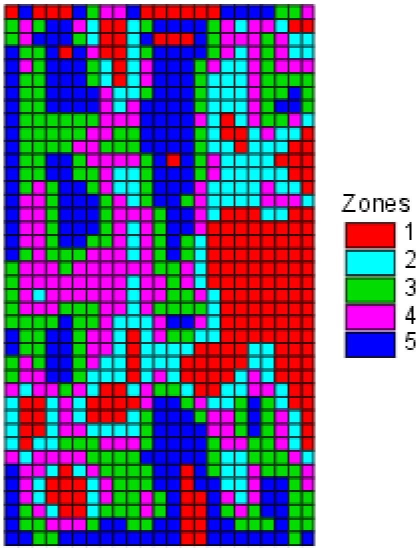

Figure 8. Management Zone Delineated using variance reduction information

The original 13 management zones identified using FPI and MPE criteria were aggregated using the difference in average yield into five zones as indicated in Table 1. The result is given in Figure 9.

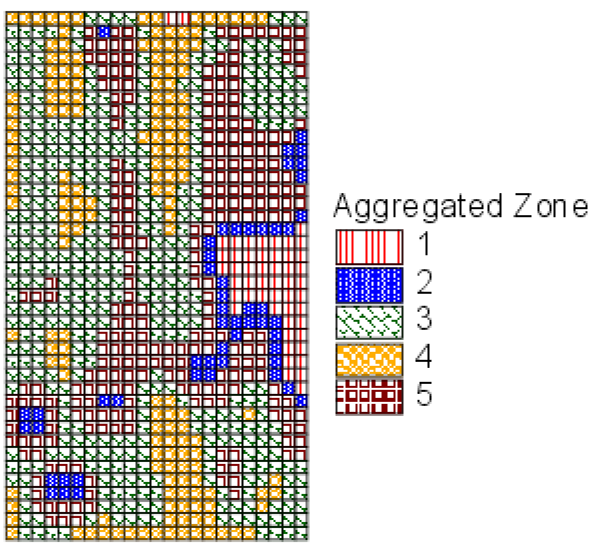

Figure 9. Management Zone Delineated By aggregating original 13 zones into 5 zones.

Table 3 shows the average yield within each of these aggregated zones. The variance reduction information shows the total within-zone yield variance: 56% of entire field for year 2000 and 70% of entire field variance for year 1998.

Table 3. Mean grain yield inside each zone
(Based on the aggregated zones)

ZONE	YIELD1998	YIELD 2000
1	1616	4363
2	1445	3113
3	1188	1863
4	1240	2233
5	1130	1486

The table has shown the requirement to increase the number of zone for 1998 to achieve the same amount of efficiency as that of the year 2000. This indicates that though we can defined a management zone to fit a condition for a particular year, we have to increase or decrease management zone based on the most important limiting factor moisture condition.

The results of the classification explained above were compared with the management zone delineated using crops and soil information. The correlation between grain yield and these factors are given in Table 4.

Table 4. PEARSON CORRELATION MATRIX FOR 2000.

	NDVI	YIELD	ABHORI	BIOM	H2O	N
NDVI	1					
YELD	0.622	1				
ABHORI	0.353	0.312	1			
BIOM	0.529	0.855	0.193	1.0		
H2O	0.500	0.458	0.744	0.313	1	
N	0.083	0.171	-0.057	0.160	-0.069	1

The correlation matrix shows the strong relationship between vegetation index and yield. For soil and crop information biomass, spring soil moisture are strongly related to the grain yield. For the delineation of management zone based on soil and crop information, we have used biomass, spring moisture and spring soil nitrogen. The result of 5 management zones delineated using soil and crop information is given in the Figure 10 below. Due to the strong relationship between crop biomass and grain yield, when compared with the yield surface it matches more that we have identified using NDVI information alone.

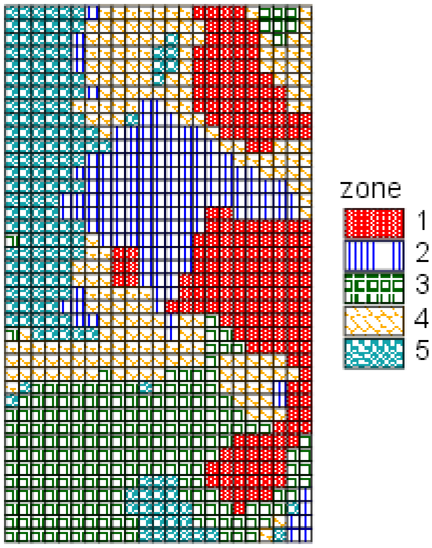

Figure 10. 5 Management Zone Delineated using crops & Soil Information (Biomass, Soil N and Soil Moisture)

The average yield inside each of the five zones delineated using soil and crop information is given in the Table 5. We have presented the results from the five zones to compare with the optimal zones i.e. five identified using the vegetation index. The total within-zone yield variance: 568925 (57% of entire field variance) which is comparable to that of zones identified based on the aggregation of original 13 zones.

Table 5. MEAN GRAIN YIELD INSIDE EACH ZONE
(Management zone based on crop & soil information)

ZONE	YIELD
1	3182
2	2558
3	1522
4	1887
5	1527

The crop response to N-fertilization rate and uniqueness of each zonederived from NDVI were checked by plotting N-uptake and grain yield against N-rate. The result is given in Figure 11 and 12 which shows that fertilizer response in each of these management zone is different, thus validating the management zone delineation process. Future work will focus on validating low-cost methods for delineating management zones for fertilization using data from this site and others.

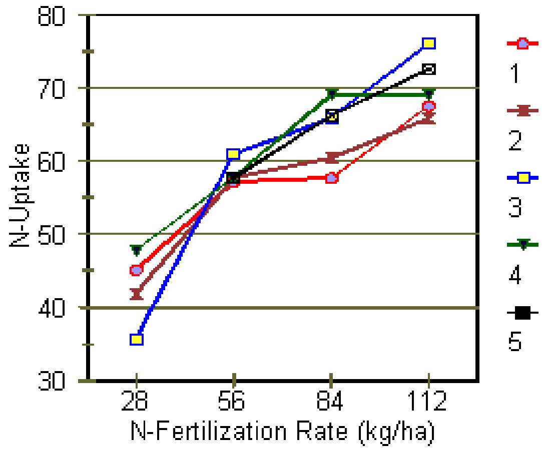

Figure 12. Response to N Fertilization inside Each Management Zone In 1997

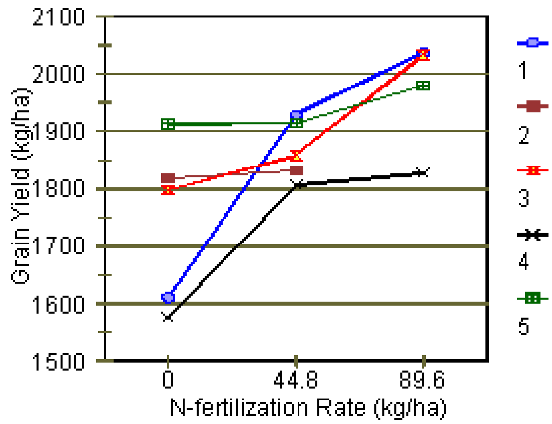

Figure 11. Response to N Fertilization inside Each Management Zone In 1999.

Conclusions

Infrared aerial photographs can provide important information for the delineation of in-field management zone. Unsupervised clustering process such as described here can help in the delineation process. In the absence of yield information index such as FPI and MPE can be used to select best number of in-field zones. A greater number of zones is generally required during years with below average precipitation (e.g.; in 1998).
This is the first step in identification of in-field management zones, which can be improved with the temporal information.

Hardware and Software Used

- GPS unit (FOR POSITION AND ELEVATION INFORMATION)
- DESKTOP COMPUTER (FOR DATA PROCESSING AND ANALYSIS)
- IDRISI SOFTWARE (FOR IMAGE ANALYSIS)
- ARCVIEW SOFTWARE (TOPOGRAPHIC, SOIL AND CROPS DATA ANALYSIS, AND FINAL ANALYSIS)
- FUZME2 SOFTWARE (FOR CLUSTERING)

Literature Cited

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.
Campbell, J.B. 1987. Introduction to remote sensing. Guilford Press, New York, 551 p.
Colwell, J.E. 1974. Vegetation canopy reflectance. Remote Sensing and Environment. 3:175-183.
deGruijter, J.J., McBratney, A.B., 1988. A modified fuzzy k means for predictive classification. In: Bock,H.H.(ed) Classification and Related Methods of Data Analysis. pp. 97-104.

Elsevier Science, Amsterdam.
Denison, R.F., R.O. Miller, D. Bryant, A. Abshahi, and W.E. Wildman. 1996. Image processing extracts more information from color infrared aerial photographs. California Agriculture. 50(3):9-13.
Fraisse, C.W., K.A. Sudduth, N.R. Kitchen, and J.J. Fridgen 1999. Use of unsupervised clustering algorithms for delineating within-field management zones. Paper No. 993043, ASAE, Toronto, Ontario, Canada.
Frazier, B.E. and Y. Cheng. 1989. Remote sensing of soils in the eastern Palouse region with Landsat thematic mapper data. Remote Sensing and Environment, 28:317-325,
Harrison, W.D., M.E. Johnson, and P.F. Biggam. 1987. Video image analysis of large scale vertical aerial photography to facilitate soil mapping. Soil Survey Techniques (W.U. Reybdd and G.W. Peterson, editors), Soil Science Society of America Special Publication 20, ASA-CSSA-SSSA, Madison, Wisconsion. Pp 1-9.
Large, E.C. 1954. Growth stages in cereals. Plant Pathology.3:128-129.
Lillesand , T.M., and R.W. Kiefer. 1994. Remote sensing and image interpretation (third edition). John Wiley and Sons, Inc., New York, 750 p.
Long, D.S., G.A. Nielson, and G.R. Carlson. 1989. Use of aerial photographs for improving layout of field research plots. Application of Agricultural Research. 4:96-100.
McConkey, B., B. McCann, F. Selles, and D. Ulrich 1999. Problems using intensive sampling of soil moisture and nitrates to predict yield variation: Inferences for site-specific farming. (unpublished paper).
McBratney, A.B., Moore, A.W., 1985. Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology 35, 165-185.
Minasny, B., McBratney, A.B., 2000. FuzME version 2.1, Australian Centre for Precision Agriculture, The University of Sydney, NSW 2006. http://www.usyd.edu.au/su/agric/acpa).
Milfred, C.J., and R.W. Kiefer. 1976. Analysis of soil variability with repetitive aerial photography. Soil Science Society of America Journal. 40:553-557.
Moran, M.S., Y. Inoue, and E.M. Barnes 1997. Opportunities and limitations for image-based remote sensing in precision crop management. Remote Sensing of Environment. 61:319-346.
Price, J.C. 1987. Calibration of satellite radiometers and the composition of vegetation indices. Remote Sensing of Environment.21:15-27.
Prince, S.D. 1990. High temporal frequency remote sensing of primary production using NOAA AVHRR. In Applications of Remote Sensing in Agriculture pp. 169-184. Butterworths, London.
Roubens, M., 1982. Fuzzy clustering algorithms and their cluster validity. European Journal of Operational Research 10, 294-301.
Sallers, P.J. 1989. Vegetation-canopy spectral reflectance and biophysical processes. In Theory and Applications of Optical Remote Sensing pp 297-335. Wiley, New York.
Sawyer, J.E. 1994. Concepts of variable rate technology with consideration of fertilizer technology. Journal of Production Agriculture, 7:195-201.
Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment. 8:127-150.
Wiegand C.L., A.J. Richardson, D.E. Escobar, and A.H. Gerbermann 1991. Vegetation indices in crop assessments. Remote Sensing of Environment.35:105-119.

Wiegand, C.L., J.D. Rhoades, D.E. Escobar, and J.H. Everitt 1994. Photographic and videographic observations for determining and mapping the response of cotton to soil salinity. Remote Sensing of Environment. 49:212-223.

Zheng, F., and H. Schreier 1988. Quantification of soil patterns and field soil fertility using spectral reflectance and digital processing of aerial photographs. Fertilizer Research. 16:15-30.