Delineation of site-specific management zones by fuzzy clustering of soil and topographic attributes: A case study of East Nile Delta, Egypt

A Saleh1 and A A Belal
Soil Sciences Dept., National Authority for Remote Sensing and Space Sciences (NARSS), Cairo, Egypt.
E-mail: ahmedms@outlook.com

Abstract. The objective of this study was to define site-specific management zones of 67.2 ha of a wheat pivot field at East of Nile Delta, Egypt for use in precision agriculture based on spatial variability of soil and topographic attributes. The field salinity was analysed by reading the apparent soil electrical conductivity (ECa) with the EM38 sensor horizontally and vertically at 432 locations. The field was sampled for soil attributes systematically with a total of 80 sampling location points. All samples were located using GPS hand held unit. Soil sampling for management zones included soil reaction pH, soil saturation percentage, organic matter, calcium carbonates content, available nitrogen, available phosphorus and available potassium. The field topographic attributes were digital elevation model (DEM), slope, profile curvature, plane curvature, compound topographic index (CTI) and power stream index (PSI). The maps of spatial variability of soil and field topographic attributes were generated using ordinary kriging geostatistical method. Principal component analysis (PCA) was used to determine the most important soil and topographic attributes for representing within-field variability. Principal component analysis of input variables indicated that EM38 horizontal readings (EM38h), soil saturation percentage and digital elevation model were more important attributes for defining field management zones. The fuzzy c-means clustering method was used to divide the field into potential management zones, fuzzy performance index (FPI) and normalized classification entropy (NCE) were used to determine the optimal cluster numbers. Measures of cluster performance indicated no advantage of dividing these fields into more than five management zones. The defined management zones not only provided a better description of the soil properties, but also can direct soil sampling design and provide valuable information for site-specific management in precision agriculture.

1. Introduction
Site-specific applications of agricultural inputs can be implemented by dividing a field into smaller Management zones that are more homogeneous in properties of interest than the field as a whole. A Management zone is defined 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 [1]. Thus, Management zones within a field may be different for different inputs and the delineation of MZs for a specific input involves only the factors directly influencing the effectiveness of that input in achieving certain goals. The number of distinct MZs within a field is a function of the natural variability within the field, the size of the field and certain management factors. The removal of excessive details in within-field variability simplifies the shapes of the zones. Thus, it reduces the equipment requirements for Variable Rate Technology (VRT) [2,3]. Management zones can be delineated based on a fast and non-destructive measurement of the apparent electrical conductivity of the soil (ECa) with the aid of electrode-based and electromagnetic methods (ECa). Factors that influence ECa include soil salinity, clay content and clay mineralogy, soil pore size and distribution,
soil moisture content and temperature [4,5,6,7]. Electromagnetic induction can be used to indirectly estimate soil properties if the contributions of the other soil properties affecting the ECa measurement are known or can be estimated. Previous studies have shown that the controlling factor in some areas is clay content [8,9,10,11,12], salinity [13,14,15] and water content [16,17,18] and can include all three factors in different parts of a study area [19,20].

2. Study area
The study area is a selected pivot (Pivot 78) in Sixth of October Company for Agricultural Projects (SOAP) which located in El-Salhia Area, East of Nile Delta, Egypt (Fig. 1). It is bounded by 31° 58’ 30” and 31° 59’ 05” longitudes and 30° 25’ 55” and 30° 26’ 30” latitudes with a total area of 154 feddan. The Sixth of October Company for Agricultural Projects have large pivots (more than 100 feddan per pivot) for wheat planting that are appropriate for practices for precision agriculture.

3. Methodology
3.1. Soil variables
The pivot field was sampled systematically with a total of 80 soil sampling location points. The sampling depth was 0.0 - 0.30 m where four sub samples were taken from the top 0.30 m of soil to create a composite sample for each sample. The analyses included soil pH (extract), organic matter (%), calcium carbonates content (%), available nitrogen (mg /kg), available phosphorus (mg /kg) and available potassium (mg /kg). the analyses were performed according to [21]. The apparent soil conductivity readings were measured using Electromagnetic Induction (EM38) device. A number of 432 EM38 horizontal and vertical survey readings were measured along 10 transects grid across the pivot area. Interpolation between sampling locations was made by ordinary Kriging geostatistical interpolation method.

3.2. Topographic variables
Topographic variables offer a large potential to characterize the within-field soil variation [22], have a direct link with pedogenetic processes [23] and closely associated with soil water availability [24]. The topographic variables were digital elevation model (DEM), slope, profile curvature, plane curvature, compound topographic index (CTI), and power stream index (PSI). Profile curvature affects the acceleration or deceleration of flow across the surface [25]. Plane curvature relates to the convergence and divergence of flow across a surface [26]. CTI is highly correlated with several soil attributes such as horizon depth, silt percentage, organic matter content, and phosphorus [27].

3.3. Factor analyses
Factor analyses were performed to determine the most soil and topographic variables influencing the management zones extraction. At first, the correlation matrix between soil and topographic variables was calculated to determine how the relationship between the variables is. Then, the principal components analyses were used as the method for factor analyses based on the correlation matrix [28]. The PCs extraction begins by finding a linear combination of variables that accounts for as much variation in the original variables as possible [29]. A varimax rotation of the solution was performed to maintain the cumulative percentage of variation explained by the extracted components [30 and 31]. The initial communalities, extracted communalities, rotated communalities, factor score coefficient matrix, anti-image covariance, and correlation matrices were calculated [32]. Kaiser-
Meyer-Olkin Measure of Sampling Adequacy was used to test the hypothesis that soil and topographic variables are unrelated and therefore unsuitable for structure detection. Cronbach's Alpha test of reliability [33] was used to measure the extraction reliability.

3.4. Fuzzy clustering

Fuzzy c-means [34 and 35] employs fuzzy partitioning such that an instance can belong to all clusters with different degrees of membership between 0 and 1. The aim of fuzzy c-means is to find cluster centroids that minimize an objective function. The fuzzy c-means clustering algorithm proposed by [36] was used for the purpose of partitioning n data observations in feature space into c-groups or clusters based on a fuzzy c-means partition. Mahalanobis distance [37] used as the metric distance and fuzzy exponent equals 1.30. The maximum number of classes was set to 10 classes. The fuzzy clustering was set to random start with 0.1 membership scatter and one trial. The fuzzy clustering process was iterated 1000 times with 0.0001 stopping criteria. The discriminant analyses of fuzzy clustering were calculated. A bootstrap jack-knifing technique was also used to estimate the fuzzy clustering with each data point temporarily removed from the data set. The bootstrap was calculated with 80% of sampled data and 50 runs. The same inputs and constrains of the fuzzy clustering process were also used for jack-knifing technique. The fuzziness performance index (FPI) [38 and 39] and normalized classification entropy (NCE) [35] were used to determine the optimal number of clusters for the pivot field. FPI measures the degree of separation between fuzzy c-partitions of X. NCE models the amount of disorganization of a fuzzy c-partition of X [37 and 39]. The optimal number of clusters for each computed index is when the index is at the minimum, representing the least membership sharing and the greatest amount of organization as a result of the clustering process [40].

4. Results and discussion

4.1. Factor analyses

The correlation matrix between the soil and topographic variables showed some positive and negative correlation between variables. As the correlation relationship between variables influences the management zones extraction and lowers the performance of the process, principal components analyses were used to make new components that account for as much variation in the original variables. The factor loading plots showed that soil saturation percent (SP), EM38 vertical signal data (EMv), and digital elevation model (DEM) variables have similar heavy loadings for the first and second principal component (PC1). The potassium content (K) and slope variables, however, have similar heavy loadings for the second and third principal components respectively. The Kaiser-Meyer-Olkin measure value (0.60) value indicates that the results of the factor analysis probably won't be very useful for management zones analyses. Bartlett's test of sphericity indicated the relation and structure between variables. Also, the Cronbach's Alpha test of reliability value (0.401) indicated the unreliability of the components extraction. According to the results, the unreliable variables were excluded and the principal components extraction was repeated. For each repeat, the unreliable variables were excluded. At the fifth extraction, the Kaiser-Meyer-Olkin measure of sampling Adequacy value (0.74) indicated that the selected principal components extraction variables have the lowest structure detection and Bartlett's test of sphericity value level indicate the usefulness of the principal components. The high communalities for the fifth extraction indicated that the extracted components represent the variables well. The Cronbach's Alpha test also reached its maximum value (0.922) at the fifth extraction. According to the results of the factor analyses, the most suitable variables for management zones extraction are EM38 horizontal signal data (EM38h), soil saturation percent (SP), and digital elevation model (DEM).
4.2. Fuzzy clustering

The management zones were clustered by the fuzzy c-means clustering technique to the selected variables (EM38h, SP, and DEM). The cluster analyses were performed to a maximum number of 10 classes. The optimal number of clusters for each computed index is representing the least membership sharing (FPI) or greatest amount of organization (NCE) as a result of the clustering process, when the index is at the minimum. The results from the two indices are plotted in Figure (2). The fuzziness performance index (FPI) and normalized classification entropy (NCE) had the same change in trend with an increase in cluster number, and the minimum fuzziness performance index (FPI) and normalized classification entropy (NCE) were obtained with 5 clusters which indicates that the sum of squares for members within a cluster is minimized while the sum of squares between members of different clusters is maximized. The fuzziness performance index (FPI) reached its minimum (0.20) at 5 classes which indicates appropriate distinct classes with membership sharing; also the normalized classification entropy (NCE) reached its minimum (0.08) at 5 classes which indicates least membership sharing and the greatest amount of organization as a result of the clustering process. The management zones were mapped Figure 3 with the corresponding voroni polygons.

Figure 2. Graphs of fuzzy performance index (FPI) and normalized classification entropy (NCE) plotted against number of clusters based on the results of fuzzy k-means classification.

5. Conclusion

The objective of this study was to define site-specific management zones of 67.2 ha of a wheat pivot field at East of Nile Delta, Egypt for use in precision agriculture based on spatial variability of soil and topographic attributes. The soil management zones were achieved in this work using geostatistics, principal component analysis, and fuzzy logic to handle the data. It was proved to be a viable procedure for the area, allowing the delineation of homogeneous and distinct regions among them with reference to soil and topographic attributes. The resulting five management zones represent a reasonable number for practical use. Soil apparent electric conductivity (EM38h), soil saturation percent (SP), and digital elevation model (DEM) are the main factors controlling the spatial variability of other soil attributes. The delineated management zones can be used to characterize spatial variability in soil properties.

Figure 3. Management zones map with corresponding voroni polygons.

6. Acknowledgement

The authors of this research acknowledge the Science and Technology Development Fund (STDF), Egypt for the financial support and also acknowledge the Sixths of October Company for Agricultural Projects for the facilitation of conducting this
References

[1] Doerge T 1998 Defining management zones for precision farming. Crop Insights 8 21
[2] Chang J, Clay D E, Carlson C G, Clay S A and Reese C 2000 Determining the impact of approaches to classify nutrient management zones In P. Robert (ed.) Proceedings of the Fifth International Conference on Precision Agriculture, Minneapolis MN, July 16-19 2000
[3] Zhang N and Taylor R 2000 Applications of a field-level geographic information system (GIS) in precision agriculture. In proceedings of the Fifth International Conference on Precision Agriculture P. C. Robert, R. H. Rust, W. E. Larsen (eds). Bloomington, Madeshon, USA.
[4] James I T, Waine T W, Brandley R I, Godwin R J and Taylor J C 2000 A comparison between traditional methods and EMI scanning for determining soil textural boundaries Agenda 2000 Warwick, paper no. 99-PA-014.
[5] Hendrickx J, Baerends B, Raza Z, Sadig M, and Chaudhry A 1992 Soil salinity assessment by electromagnetic induction of irrigated land SSSA J 56 1933-1941
[6] McNeill J D 1992 Rapid, accurate mapping of soil salinity by electromagnetic ground conductivity meters, Advances in Measurements of Soil Physics Properties: Bringing Theory into Practice, Soil Science Society of America Special Publication 30, American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, Wis-consin 201–229
[7] Rhoades J D, Chanduvi F and Lesch S M 1999 Soil salinity assessment: Methods and interpretation of electrical conductivity measurements, FAO Irrigation and Drainage Paper #57, FAO, Rome, Italy, 150
[8] Durlesser H 1999 Bestimmung der variation bodenphysikalischer parameter in raum und zeit mit elektromagnetischen induktionsverfahren. (in German) FAM-Bericht 35 121 Technische universität münchen Eijkelkamp Agrisearch Equipment 2003 Operating instructions, EC-probe set for soil conductivity measurements.
[9] Cook P G and Walker G R 1992 Depth profiles of electrical conductivity from linear combination of electromagnetic induction measurements Soil Science American Journal 56 1015-1022
[10] Triantafilis J and Lesch S M 2005 Mapping clay content variation using electromagnetic induction techniques Computers and Electronics in Agriculture 46 203–237
[11] Hedley C B, Yule I Y, Eastwood C R, Shepherd T G and Arnold G 2004 Rapid identification of soil textural and management zones using electromagnetic induction sensing of soils Australian journal of soil research 42 389-400
[12] Dalgaard M, Have H and Nehmdahl H 2001 Soil clay mapping by measurement of electromagnetic conductivity In: Proceedings of third European conference on precision agriculture, Montpellier, France 367-372
[13] Rhoades J D, Manteghi N A, Shouse P J, and Alves W J 1989 Soil electrical conductivity and soil salinity: New formula-tions and calibrations Soil Science Society of America Journal 53 433–439
[14] Lesch S M, D J Strauss and Rhoades J D 1995a Spatial prediction of soil salinity using electromagnetic induction techniques: 1. Statistical prediction models: A comparison of multiple linear regression and cokriging Water Resources Research 31 373–386
[15] Lesch S M, Strauss D J and Rhoades J D 1995b Spatial prediction of soil salinity using electromagnetic induction techniques: 2. An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation Water Resources Research 31 387–398
[16] Kachanoski R G, De Jong E and Van Wenesbeek I J 1990 Field scale patterns of soil water storage from non-contacting measurements of bulk electrical conductivity Canadian Journal of Soil Science 70 537–541
[17] Kachanoski R G, Gregorich E G and Van Wenesbeek I J 1988 Estimating spatial variations of soil water content using noncontacting electromagnetic inductive methods Canadian Journal of Soil Science 68 715–722
[18] Sheets K R and Hendrickx J M H 1995 Noninvasive soil water content measurement using electromagnetic induction. Water Resources Research 31 2401–2409
[19] Paine J G, Goldsmith R S and Scanlon B R 1998 Electrical conductivity and gamma-ray response to clay, water and chloride content in fissured sediments, Trans-Pecos Texas Environmental and Engineering Geoscience 4 225–239
[20] Scanlon B R, Paine J G and Goldsmith R S 1999 Evaluation of electromagnetic induction as a reconnaissance technique to characterize unsaturated flow in an arid setting Groundwater 37 296-304
[21] Rowell D L 1995 Soil Science Methods and Applications: Library of Congress Cataloging in Publication data. New York, NY 10158, USA
[22] Franzen D W, Hopkins D H, Sweeney M D, Ulmer M K and Havlvorson A D 2002 Evaluation of soil
survey scale for zone development of site-specific Nitrogen management. *Agron. J.* 381-389

[23] Wilson J P and Gallant J C editors 2000 Terrain Analysis: Principles and Applications. wiley and sons.

[24] Fraisse C W, Sudduth K A, Kitchen N R 2001 Delineation of site-specific management zones by unsupervised classification of topographic attributes and soil electrical conductivity *Trans ASAE* 44 155-166

[25] Zevenbergen L W, Thorne C R 1987 Quantitative analysis of land surface topography *Earth Surface Processes and Landforms* 12 12-56

[26] Moore I D, Gessler P E, Nielsen G A and Petersen G A 1993 Terrain attributes: estimation methods and scale effects. In Modeling Change in Environmental Systems, edited by Jakeman A J, Beck M B and McAleer M (London: Wiley) 189 - 214

[27] Kline P 1994 An Easy Guide to Factor Analysis. London: Routledge

[28] Gorsuch R L 1983 Factor Analysis 2nd ed. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

[29] Kim J and Mueller C W 1978a Factor Analysis: What It Is and How To Do It. Beverly Hills and London: Sage Publications

[30] Kim J, and Mueller C W 1978b Factor Analysis: Statistical Methods and Practical Issues. Beverly Hills and London: Sage Publications

[31] Harman H H 1976 Modern Factor Analysis, 3rd ed. Chicago: University of Chicago Press

[32] Cronbach L J 1951 Coefficient alpha and the internal structure of tests. *Psychometrika* 16 297-334

[33] Dunn J C 1973 A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters *Journal of Cybernetics* 3 32-57

[34] De Grujiter J J and McBratney A B 1988 A modified fuzzy k-means for predictive classification. In: Bock, H.H. (Ed) Classification and Related Methods of Data Analysis Elsevier Science 97–104

[35] Bezdek J C 1981 Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York

[36] McLachlan G J 1992 Discriminant Analysis and Statistical Pattern Recognition, Wiley Interscience

[37] Odeh I O A, McBratney A B, Chittleborough D J 1992 Soil pattern recognition with fuzzy-c-means: application to classification and soil-landscape interrelationships *Soil Sci. Soc. Am. J.* 56:505-516

[38] Boydell B, McBratney A B 2002 Identifying Potential within-Field Management Zones from Cotton-Yield Estimates *Precision Agriculture* 3 9-23

[39] Lark R M and Stafford J V 1997 Classification as a first step in the interpretation of temporal and spatial variation of crop yield *Ann. Appl. Biol* 130 111-121

[40] Fridgen J J, Kitchen N R and Sudduth K A 2000 Variability of Soil and Landscape Attributes within Sub-field Management Zones. In: Robert P C, Rust R H, Larson W E (Eds.) Proceedings of the 5th International Conference on Precision Agriculture and Other Source Management. ASA, CSSA, SSSA, Madison, WI, USA.