AN APPLICATION OF HYBRID CLUSTERING AND NEURAL BASED PREDICTION MODELLING FOR DELINEATION OF MANAGEMENT ZONES

[1]Babankumar S. Bansod
[1] Agrionics Department
CSIR-CSIO, Chandigarh
[2]OP Pandey
[2]Thapar University
Patiala

Abstract: Starting from descriptive data on crop yield and various other properties, the aim of this study is to reveal the trends on soil behaviour, such as crop yield. This study has been carried out by developing web application that uses a well known technique- Cluster Analysis. The cluster analysis revealed linkages between soil classes for the same field as well as between different fields, which can be partly assigned to crops rotation and determination of variable soil input rates.

A hybrid clustering algorithm has been developed taking into account the traits of two clustering technologies: i) Hierarchical clustering, ii) K-means clustering. This hybrid clustering algorithm is applied to sensor- gathered data about soil and analysed, resulting in the formation of well delineated management zones based on various properties of soil, such as, EC_a , crop yield, etc.

One of the purposes of the study was to identify the main factors affecting the crop yield and the results obtained were validated with existing techniques. To accomplish this purpose, geo-referenced soil information has been examined. Also, based on this data, statistical method has been used to classify and characterize the soil behaviour. This is done using a prediction model, developed to predict the unknown behaviour of clusters based on the known behaviour of other clusters. In predictive modeling, data has been collected for the relevant predictors, a statistical model has been formulated, predictions were made and the model can be validated (or revised) as additional data becomes available. The model used in the web application has been formed taking into account neural network based minimum hamming distance criterion.

Keywords: precision farming, management zones (MZ), hybrid clustering, prediction modelling, neural network

I. INTRODUCTION

A. Precision farming

Precision agriculture, firstly advanced by American agriculturalists at the beginning of 1990s, is a modern agricultural production mode and a technical system synthetically applying modern high and new technologies, such as 3S (RS, GIS and GPS), artificial intelligence etc., to obtain high yield, high quality and high efficiency. Variable-rate fertilization is the key technique and means to realize precision agriculture, but traditional variable-rate fertilization is mainly based on grid sampling method to obtain information, which is expensive and not practical. In recent years, many scholars have put their efforts on the researches to divide a field into several relatively homogeneous sub-zones on the basis of the spatial variability and location of soil nutrient, namely, to delineate soil nutrient management zones. Scientific and rational technique for delineating soil nutrient management zones is an efficient means to carry out variable-rate fertilization for precision agriculture and becomes a hotspot of precision agriculture research abroad and in home.(as in [5])

Efficient use of agro-chemicals is beneficial for farmers as well as for the environment. Spatial and temporal optimization of farm management will lead to an increase in the productivity or reduction in the amount of agro-chemicals. This type of management is referred to as Precision Agriculture. Traditional management implicitly considers any field to be a homogeneous unit for management: fertilization, tillage and crop protection measures, for example, are not varied within a single field. The question for management is what to do when. Because of the variability within the field, this implies inefficient use of resources. Precision agriculture defines different management practices to be applied within single, variable fields, potentially reducing costs and limiting adverse environmental side effects. The question is not only what and when but also where. Many tools for management and analysis of spatial variable fields have been developed. (as in [11])

Optimal/precision farming may be used to improve a field or a farm management from several perspectives:

- agronomical perspective: adjustment of cultural practices to take into account the real needs of the crop (e.g., better fertilization management)
- technical perspective: better time management at the farm level (e.g. planning of agricultural activity)
- environmental perspective: reduction of detrimental agricultural impacts (better estimation of crop nitrogen needs implying limitation of nitrogen run-off)
- economical perspective: increase of the output and/or reduction of the input, increase of efficiency (e.g., lower cost of nitrogen fertilization practice)

B. Management zones

Management zones (MZ) are defined as sub-regions of a field that has a relatively homogeneous combination of yield-limiting factors, for which a single rate of a specific crop input is appropriate to attain maximum efficiency of farm inputs .Besides representing areas of equal production potential, within-field management zones have many other uses. Several studies have indicated that homogenous management zones could be used as an alternative to grid soil sampling and to develop nutrient maps for variable rate fertilizer application .Spatially coherent areas within fields have also proved useful in relating yield to soil and topographic parameters for crop-modelling evaluation (as in [5]).

Based on descriptive information about the crop yield, the aim of the developed web application is to identify trends in crops production, as well as linkages between productions of different crops. The web application, thus, works on two algorithms:

a) Hybrid clustering, using the concept of Hierarchical clustering and K-means clustering to form management zones.

b) Prediction modelling, using the concept of Hamming distance to predict the behaviour of zones, thus formed.

C. Clustering

Data clustering is a common technique for statistical data analysis, which is used in many fields, including data mining and bioinformatics (as in [7]).Cluster analysis techniques have been applied to the aim of detecting any significant spatial variation in the skill of soil behaviour prediction models. The subsets detected by the clustering algorithm were remarkably persistent in the events in exam, and forecast quality appears to depend on the spatial variability represented by clusters.
In order to identify the most significant patterns in the wide number of available observation points, suitable criteria to gather them around representative modes have been borrowed from clustering analysis techniques. Several hierarchical agglomerative clustering techniques have been tested (as in [4]) in various applications. The hybrid clustering algorithm has been selected due to its good performances as compared to the individual algorithms when applied separately. A key issue in this application of cluster analysis is the use of the appropriate input dataset to relate different field regimes to specific geographical areas. Thus, a two dimensional phase space has been adopted where dimensions include normalized space coordinates (latitude and longitude) and various soil properties (like EC_a, pH, soil moisture, etc.) were measured using sensors such as Geonics, Pogo-hana, etc. Also crop yield was measured at different grids within the field.

D. Prediction modelling

The developed web application predicts the behaviour of created zones using prediction modelling. The accuracy of all predictions made using numerical modelling is strictly limited by the natural variability of geologic materials. In this web application, an attempt is made to quantify this accuracy through the straightforward application of hamming distance (as in [9]).

Thus using this web application, a user will be able to:

a) Delineate the boundaries of a given field based on various soil properties, thereby forming various management zones.

b) Predict the behaviour of the management zones thus formed based on the previously known behaviour of some management zones.

II. MATERIALS AND METHODOLOGY

After logging in to the web application(as in Fig.1), the user can specify the name of the database and the name of the table that stores the properties measured by sensors, such as Geonics or pogo-hanna in the present case, by selecting the appropriate option from the available menu to form the clusters. The clusters will be formed by specifying the name of the column based on desired selection of soil attribute (as in Fig.2 and Fig.3).

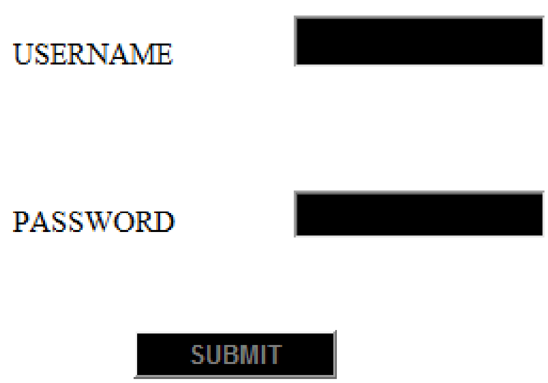


Fig.1. Login screen

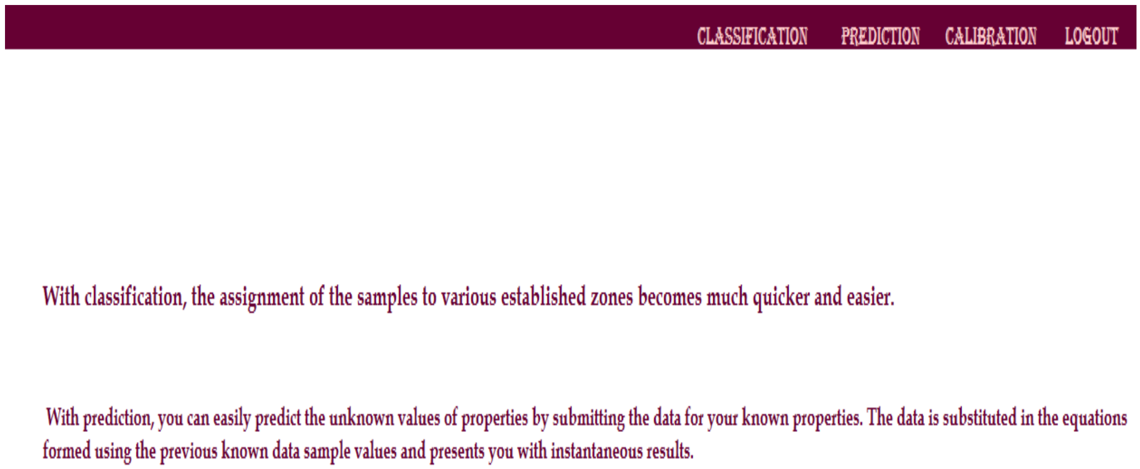


Fig.2. Classification option using hybrid clustering

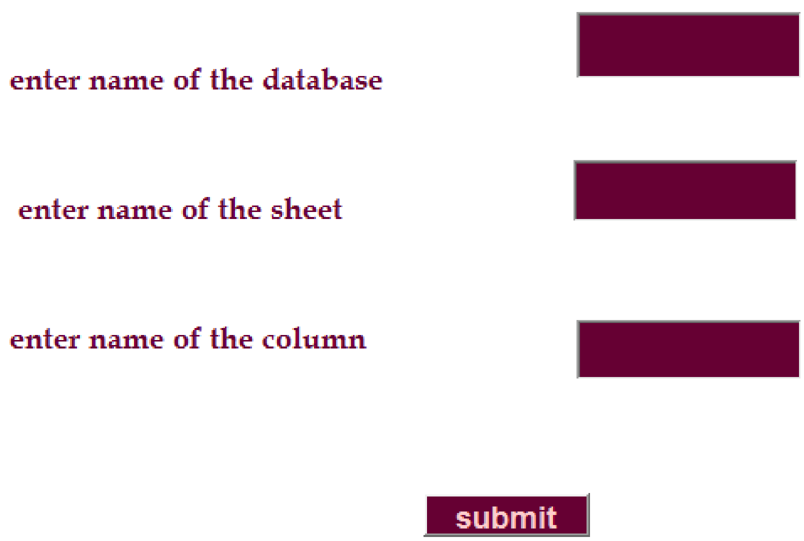


Fig.3. Database information

A. Clustering algorithm:

a) Consider each value v_i belongs to a different cluster c_i.

$$v_i \in c_i \quad (1)$$

b) Find mean m of the values under consideration.

$$m = \frac{\sum_{i=1}^{n} v_i}{n} \quad (2)$$

c) Form ranges r_a-r_b, where $1<a<n$ and $1<b<n$ based on the distance from the mean using iteration till the mean m_i of the lowest range is greater that the min value v_1 and the mean m_j of the greatest range is less than the maximum value m_n.

d) The no. of iterations gives the optimal no. of clusters.

e) If the value v_i lies within range r_a-r_b, the point lies in cluster c_a, where a belongs to i.

The cluster formation is thus done using the web application (as in Fig.4).

ID	bioyield	Cluster
1	836	8
2	3806.87	4
3	3414.32	4
4	3810.92	4
5	3607.22	4
6	3607.22	4
7	3414.32	4
8	3810.92	4
9	836	8
10	3806.87	4
11	3806.87	4
12	836	8
13	3414.32	4
14	3810.92	4
15	3607.22	4
16	5217.93	8

Fig.4. An instance of cluster formation using bio-yield

B. Prediction modelling rule:

Once the management zones have been found, the next task is to predict the properties of some management zones based on the prior information about other management zones. This is done using a prediction model with hamming distance as the basis. The hamming distance concept is derived from the firing rule of a neuron in a neural network.

The rule is stated as: Take a collection of known patterns for a specific cluster, some of which represent the properties it shows (1 value set) and other which does not (0 value set). Then the patterns not in the collection have a value 1 if, on comparison , they have more input elements in common with the 'nearest' pattern in the 1-value set than with the 'nearest' pattern in the 0-value set. If there is a tie, then the pattern remains in the undefined state. For example, a 3-input pattern shows a property with 1 output when the pattern is 111 or 101 and to output 0 when the input is 000 or 001. Then, before applying the rule, the truth table is

X1:		0	0	0	0	1	1	1	1
X2:		0	0	1	1	0	0	1	1
X3:		0	1	0	1	0	1	0	1
OUT:		0	0	0/1	0/1	0/1	1	0/1	1

Table1. Truth table before applying firing rule

As an example of the way the rule is applied, take the pattern 010. It differs from 000 in 1 element, from 001 in 2 elements, from 101 in 3 elements and from 111 in 2 elements. Therefore, the 'nearest' pattern is 000 which belongs in the 0-value set. Thus the rule requires that the output should be 0 when the input is 010. On the other hand, 011 is equally distant from two taught patterns that have different outputs and thus the output stays undefined (0/1).

By applying the rule in every column the following truth table is obtained;

X1:		0	0	0	0	1	1	1	1
X2:		0	0	1	1	0	0	1	1
X3:		0	1	0	1	0	1	0	1
OUT:		0	0	0	0/1	0/1	1	1	1

Table2. Truth table after applying firing rule

Applying same rule in the web application for the field points under consideration, an input pattern was formed using prior known facts about the soil properties and the unknown clusters were categorised according to their hamming distance based on the input patterns of same soil properties as that of the known ones (refer Fig.5)

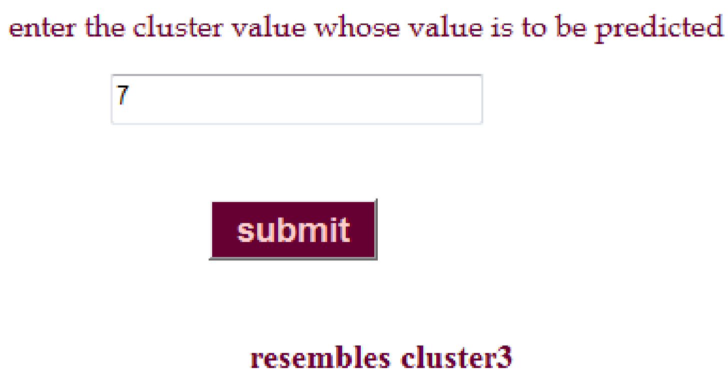


Fig.5. Prediction modelling using hamming distance

C. EXPERIMENTAL DATA:

The field survey was conducted on three fields situated in Punjab Agricultural University, Ludhiana, (INDIA) after harvesting different varieties of paddy fields during 2010 summers.

Field1 was experimental field located at farm and machinery experimental site of PAU. It was applied with variable input nitrogen rate and in a specific grid pattern as given below, having three varieties of crop planted in different sections. An arrangement of Field1 was divided into 3 different parts (part1, part2 and part2) having grids of 3by5 totalling 15 grid cells in each part, each cell having dimension of 6by5 metre, a one metre width row was left between two grids. The data collection using sensor was carried out at the vertices of each cell. This resulted in to collection of 24 points from each grid i.e. 72 points in total from field 1 (refer Fig.6).

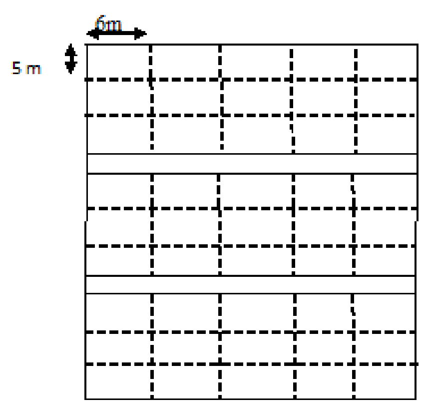


Fig.6. Field 1- Grid Outline

Field2 was divided in to 4by5 grid, each cell having dimensions of 9m by 9m giving 30 sampling points which were measured at the vertices of each grid cell (refer Fig.7).

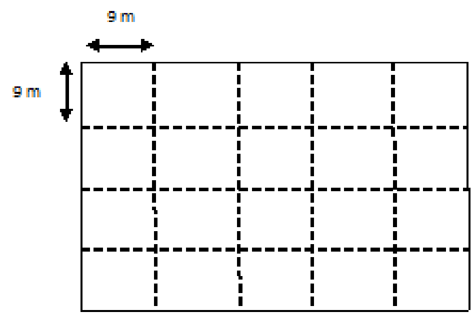


Fig.7. Field 2- Grid outline

Field3 was divided into 11by1 grid, each cell having dimension of 9m by 9m providing 24 points for the measurement. (refer Fig.8)

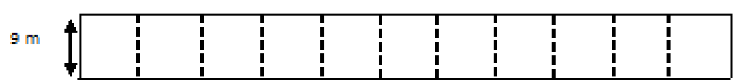


Fig.8. Field 3- Grid outline

The data was collected using Geonics EM38MK2 and Trimble DGPS after following appropriate instrument calibration procedure by referring manual. Data logger for Geonics EM38MK2 was implemented using laptop, which was also connected with Trimble GPS for geo-referencing. On-the-go geo-referenced apparent electrical conductivity (EC_a) data for each field was collected using lab developed trolley which is free from EM interferences and is made up of low induction number material i.e. wooden and data logger system having DGPS was attached to same set-up in order to have real time geo-referenced EC_a logging.

D. Soil Sampling Scheme using ESAP

After collection of field data, an EC_a directed soil sampling was done using software ESAP. It accepted the .XYZ format file. Signal de-correlation was performed over the data and outliers were seen and removed using signal validation. After signal validation, configuration of data was accepted. Now the Response Surface Sample design was created. Sample Design was created for different adjustment factor and different number of samples to give the Optimality criteria < 1.3. Sample design was saved in txt format and .jpg format. Now the EC_a directed Soil sample design was generated for each field. Twelve sampling points for each part of Field1 was designed and twenty sampling points for filed2 was designed.

E. Soil Sample Collection

The sampling locations that are obtained from ESAP, were converted to GPS compatible file so that the locations can be traced in the field.

The sample design obtained was converted to GPS waypoint file using GPS pathfinder software. Now the waypoint file was sent to GPS device through Data transfer option of the GPS pathfinder. Then soil samples were dried out for a 12-16 hours in normal air by spreading them on clean papers in lab. After that soil samples were sent to soil testing laboratory of PAU to perform lab analysis in order to find out soil properties.

This data was then given as input to the developed web application in order to analyse the cluster formation and predict the properties of the management zones thus formed

III. RESULTS AND DISCUSSION

A web application has been developed that forms management zones based on the technique of hybrid clustering, that takes into account the tree-like structure of hierarchical clustering, and the distance measure from the mean of given values as the basis of cluster formation, inheriting this trait from K-means clustering.

The management zones thus formed are then passed through a prediction model wherein their properties are predicted. The various management zones can then be given similar treatment in terms of input rate, crop varieties, etc, thereby increasing the crop yield and reducing the effort, proving as a step to precision farming.

IV. VERIFICATION:

After the formation of clusters based on Geonics measured ECa and crop yield data using the developed web application, it was noticed that they exhibit a similarity in terms of field grids belonging to the same cluster (as shown in Fig.9, where the numbers mentioned in the legend represents the cluster number as produced by the web application). Thus, the task of predicting the crop yield behaviour from the sensor measured conductivity becomes easier by the use of this web application and an important finding of the geonics measured conductivity being a surrogate measure to estimate the crop yield has been demonstrated.

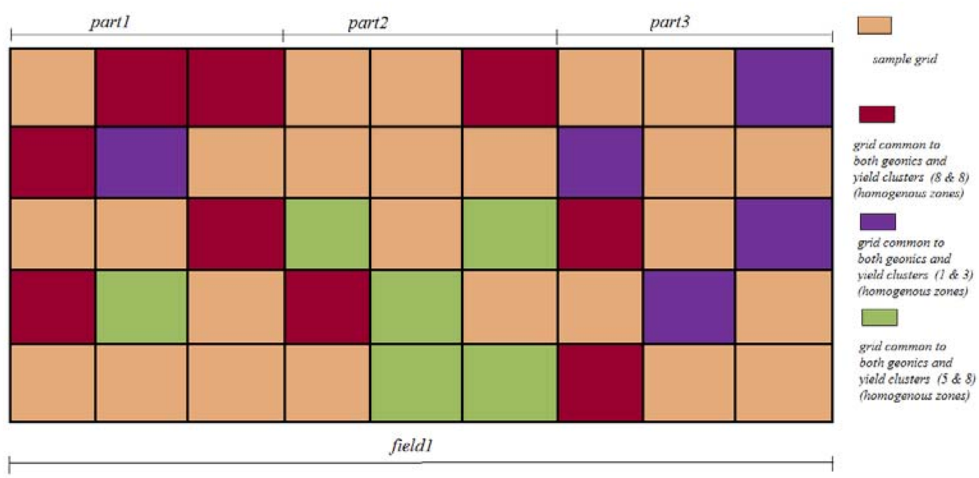


Fig 9. Common clusters for Geonics data and yield (using the web application)

This observation is in conformance to the patterns observed when the clusters are formed using established software Unscrambler X. When the experimental data was fed to this software, the clusters formed using crop yield exhibited similarity to the clusters formed using Geonics measured conductivity (refer Fig.10).

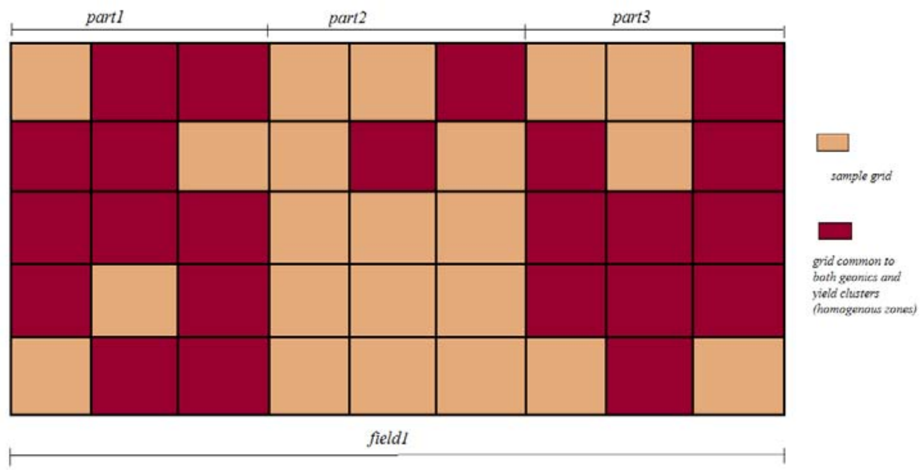
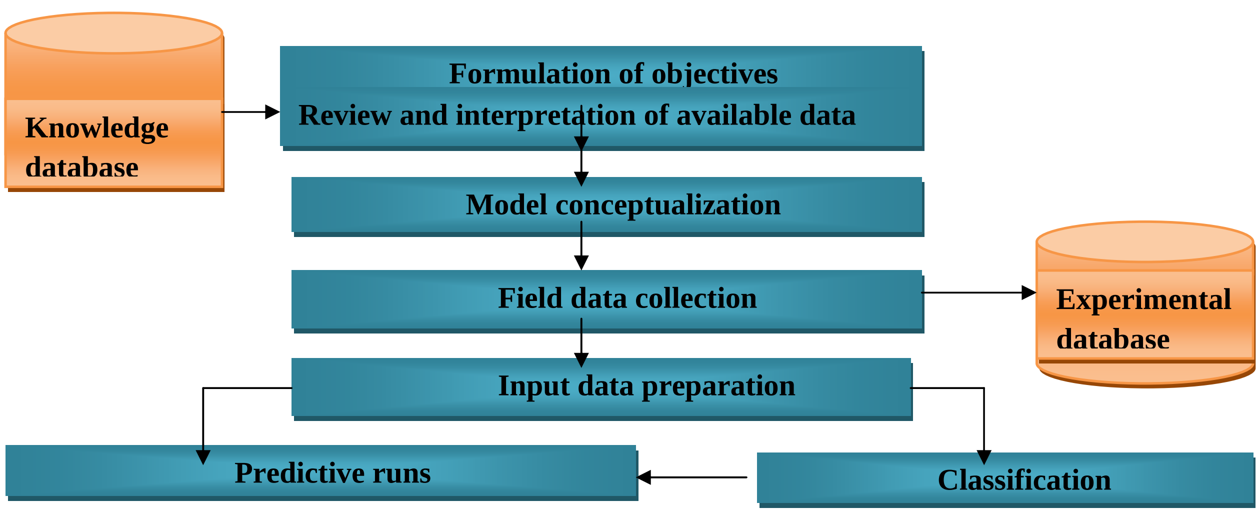


Fig.10. Common clusters for Geonics data and yield (using Unscrambler X)

V. CONCLUSION

The entire process of the development of the web application including the development of clusters based management zones and the hamming distance based prediction model can be summarized as:



Thus, the developed web application makes the work of farming easier and intelligent in a way that it enables the user to create customized management zones based on the properties of the soil and also enable the user to predict the behaviour of the soil samples belonging to a particular zone, thereby giving the probability of the tendency of a region to behave in a specific manner.

Acknowledgement

The authors express their deep sense of gratitude towards Dr. Pawan Kapur, Director CSIO, for the guidance and encouragement provided from time to time. The authors also acknowledge Ms. Shruti Trehan, trainee project, for her vital contribution in the present work.

REFERENCES

[1] Xiang Li,, Yuchun Pan, Chunjiang Zhang, Liangyun Liu, Jindi Wang, A new algorithm on delineation of management zone, Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International, 2005.

[2] Hyung Gi Byun, Krishna C. Persaud, Soad Mohialdin Khaffaf, Philip J. Hobbs, Tom H. Misselbrook , Application of unsupervised clustering methods to the assessment of malodour in agriculture using an array of conducting polymer odour sensors Original Research Article, Computers and Electronics in Agriculture, Volume 17, Pages 233-247, Issue 2, May 1997.

[3] Stefano Serafin, Alessio Bertò, Dino Zardi , Application of cluster analysis for the hydrogeochemical factors of salinegroundwater in Kimje, Korea, Geosciences Journal, Vol. 7, No. 4, p. 313_322, December 2003.

[4] Stefano Serafin, Alessio Bertò, Dino Zardi, Application of cluster analysis techniques to the verification of quantitative precipitation forecasts

[5] Qiang Fu, Zilong Wang and Qiuxiang Jiang, Delineating soil nutrient management zones based on fuzzy clustering optimized, Mathematical and computer modelling, Volume 51, Issues 11-12, June 2010, Pages 1299-1305,2009.

[6] Yan Li,Zhou Shi, Ci-fang Wu, Hong-yi Li, and Feng Li, Determination of potential management zones from soil electrical conductivity, yield and crop data, Journal of Zhejiang University Science,2008.

[7] Jäntschi L, Sorana D. Bolboacă, Carmen E. Stoenoiu, National trends on agricultural crops production:cluster analysis, Bulletin usamv-cn, 63, 2007.

[8] Antonio Mallarino, Associate Professor David Wittry, Research Associate epartment of Agronomy, Iowa State University, Management zones soil sampling: a better alternative to grid and soil type sampling, 2001.

[9] T.D. Wiles, Reliability of numerical modelling predictions ,International Journal of Rock Mechanics & Mining Sciences 43 ,454–472, 2006.

[10] Glen C. Rains,Assistant Professor Biological and Agricultural Engineering Department, , Daniel L. Thomas, Professor Biological and Agricultural Engineering Department , Soil-Sampling Issues for Precision Management of Crop Production, 229-391-3735.

[11] H. W. G. Booltink, , B. J. van Alphen, W. D. Batchelor, J. O. Paz, J. J. Stoorvogel and R. Vargas ,Tools for optimizing management of spatially-variable fields, Agricultural Systems, vol. 70, issue 2-3, pages 445-476, 2001.

[12] M. Reyniers, K. Maertens, E. Vrindts, J. De Baerdemaeker , Yield variability related to landscape properties of a loamy soil in central Belgium , Original Research Article Soil and Tillage Research, Volume 88, Pages 262-273, Issues 1-2, July 2006.