Data Fusion of Proximal Soil Sensing and Remote Crop Sensing for the Delineation of Management Zones in Arable Crop Precision Farming

Xanthoula Eirini Pantazi[1], Dimitrios Moshou[2], Abdul Mounem Mouazen[3], Thomas Alexandridis[2], Boyan Kuang[4]

[1]School of Agriculture, Aristotle University of Thessaloniki, Thessaloniki, Greece, e-mail: renepantazi@gmail.com
[1]School of Agriculture, Aristotle University of Thessaloniki, Thessaloniki, Greece
[3]Environmental Technology and Science, Cranfield University, Bedfordshire, United Kingdom, e-mail: a.mouazen@cranfield.ac.uk
[4]Environmental Technology and Science, Cranfield University, Bedfordshire, United Kingdom

Abstract. The widespread application of precision agriculture has triggered the expansion of tools for data collection and geo referencing of productivity, soil and crop properties. The correct data fusion of soil and crop parameters is a complex problem due to the abundance of inter-correlated parameters which necessitates the use of computational intelligence techniques. This paper proposes the combination of common statistical approaches with Self Organizing Clustering for delineating management zones (MZ). By this, the management of the field related to the application of inputs is becoming more accurate since the relations of the soil and crop parameters are indicated in a more precise way.

Keywords: Self-Organizing Maps, k-means, satellite remote sensing, proximal soil sensing, clustering

1 Introduction

Precision agriculture is oriented to field management taking into account its spatio-temporal variability. Its extensive use has enabled the development of tools which are capable of collecting data about soil and crop status, productivity and geolocation of these properties. The quantity of generated data demands the use of information technology in order to derive decisions concerning the management of production based on crop variability. The most widely used approach to manage the variability of fields concerns the use of MZ. Each zone is treated with the suitable level of inputs (soil tillage, seed rate, fertilizer rate, crop protection). The term of 'MZ' in a field represents a sub-region inside the field that exhibits a relatively homogeneous grouping of yield-limiting factors, concerning the treatment regime of using single rate for this zone. The MZ are defined based on soil and yield measurements, probably over a period of years (Fraisse et al., 2001). Soil information can be

Copyright © 2015 for this paper by its authors. Copying permitted for private and academic purposes.
Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2015), Kavala, Greece, 17-20 September, 2015.

effectively utilized to create 'stable' MZ which remain unaltered per field. The proper selection of parameters is regarded as a complicated task owning to the great amount of inter-correlated parameters. This leads to a nonlinear problem which can be tackled with nonlinear statistical methods and computational intelligence approaches. An improved characterization of internal variation of soil properties gives the ability to delineate MZ which reflect in a better way their true variation. Traditional soil sampling and laboratory analysis is currently not cost effective. Researches that have been recently conducted; have utilized various sensors for single soil chemical and physical attributes measurements aiming not only to decrease expenses but also to improve MZ delineation. Nevertheless, the soil-water-crop system is regarded as difficult to be characterized properly by using single property sensors (Adamchuk et al., 2004). Studies that have been lab-based, have demonstrated that the spectra of soil reflectance that originate from visible and near infra-red (vis-NIR) ranges can give direct and proxy estimations of various yield-limiting factors (Kuang et al., 2012) This success triggered the research into mobile vis-NIR sensors which would be capable of collecting soil reflectance data in situ (Shibusawa et al., 2001, Christy, 2008; Mouazen et al., 2005). These sensors are able to provide data of high resolution on soil. Prediction models have been formed by associating reflectance spectra with soil samples tested in laboratory which were obtained from the survey. These prediction models can provide local prediction maps of specific soil properties (Kuang & Mouazen, 2011). Remote sensing of vegetation has been used in Yatsenko et al. (2003) in order to estimate chlorophyll concentration from spectral data. Multi-sensor fusion is an approach that attempts to minimize the uncertainty of an estimated variable through combining data from sensors that provide observations from the entity or the phenomenon that is characterized by the mentioned variable (Boginski et al., 2012).

Data fusion of soil and crop data can be utilized for defining MZ (Taylor et al., 2003), because the data are gathered into clusters owning to similar affects between soil and crop data production mechanisms. The clusters can also formulate a starting point to discover the reasons that bring up yield variability (Reyniers, 2003).

In this study, the k-means algorithm is compared with the Self Organizing Map for delineating MZ. Further, a hybrid SOM algorithm is presented which forms clusters in combination with k-means. The hybrid SOM algorithm and k-means are compared in terms of cluster separation and MZ formation based on data fusion of Normalized Difference Vegetation Index (NDVI) and soil parameters.

2 Materials and methods

Normalized different vegetation index (NDVI) was utilized in order to the calculate crop cover and it was based on images taken by satellite which were taken two times: the first on the 2nd May and the second on the 3rd June of 2013. These satellite images were produced by Disaster Monitoring Constellation II (DMCII) for the Horns End field in the UK.

The processing workflow chain for crop NDVI is based on post-processed L1R or L1T (ortho-rectified imagery). In-band reflectance calibration was performed to obtain surface reflectance using ArcGIS. NDVI was calculated using the equation: NDVI = (NIR-R)/(NIR+R), where NIR and R is the is reflectance in the near-infrared and red bands, respectively. NDVI data were resampled to a 5mX5m grid resulting in 8798 values. A combine harvester mounted sensor was responsible for collecting yield data.

The yield was interpolated at the same 5mX5m grid as the NDVI, resulting in 8798 values. After the harvest of 2013, a spectral reflectance study utilizing the on-line vis-NIR sensor platform (Mouazen, 2006) was conducted. It comprised of an AgroSpec mobile, vis-NIR spectrophotometer of fibre type (Tec5 Technology for Spectroscopy, Germany) that covered a 305-2200 nm. 60 soil samples were gathered from the low side bottom of the trench that was opened by the subsoiler to demonstrate lab-tested levels of specific yield-limiting properties i.e. pH, phosphorus (P), potassium (K), calcium (Ca), Magnesium (Mg), organic carbon (OC), moisture content (MC), cation exchange capacity (CEC), total nitrogen (TN). Partial least squares (PLS) regression analysis was applied to soil reflectance spectra and chemical analysis values aiming to develop soil property prediction models. In order to provide point predictions, every model was fed to the on-line survey data. The creation of suitable variograms was enabled by geostatistical analysis of the prediction results. These variograms were used to give the prediction maps through interpolation by kriging. Yield data which were collected during previous harvesting periods in 2011 and 2012 was subjected to interpolation by Inverse Distance Weighting (IDW) aiming to deliver a further map layer which was capable of indicating past field fertility variation. All interpolated map layers, which were produced from the data that were collected from yield-limiting soil properties, were fused with interpolated maps of NDVI which indicated crop cover and historical yield data from years 2011 and 2012. MZ delineation by using k-means and Self Organizing Maps were performed.

3 Results

3.1 Data Fusion by Clustering with k- means

The point coordinates and property values of soil parameters, NDVI and historical yields were inserted in a spreadsheet matrix for every experimental field and then imported into Matlab software. Clustering was achieved by using the k-means clustering algorithm (Hartigan and Wong, 1979), which utilizes the unscaled, squared Euclidean distances, so as to calculate the distance., A normalization process was followed in order to avoid that a property with large values will prevail over the clustering. Normalization consisted of mean centering, followed by division with the standard deviation of the samples. This normalization was performed in order to have zero mean data which are scaled between -1 and 1. The clustering procedure enables the data fusion from numerous properties. It delineates similarity areas by putting them in the same class. Firstly, the best number of classes was determined by utilizing the gap criterion (Tibshirani et al., 2001). As regards Horn's End, the clusters were two and this was calculated by utilizing the "evalclusters" command in Matlab 2013b. This result corresponds to normalized attributes, where mean is centered and standard deviation equals to unity. In the case of non-normalized features the gap criterion is maximized for 8 clusters. The values of the GAP criterion referring to different numbers of clusters are shown in Fig.2. The result is the same when utilizing the NDVI with historic yields and soil parameters of the years 2011 and 2012 and when using only historic yields with soil parameters. Each input spreadsheet point was given an integer to show membership of a class. The acquired clusters by repeating the k-means algorithm between 2 and 7 clusters brought up the results that are demonstrated in Figure 1.

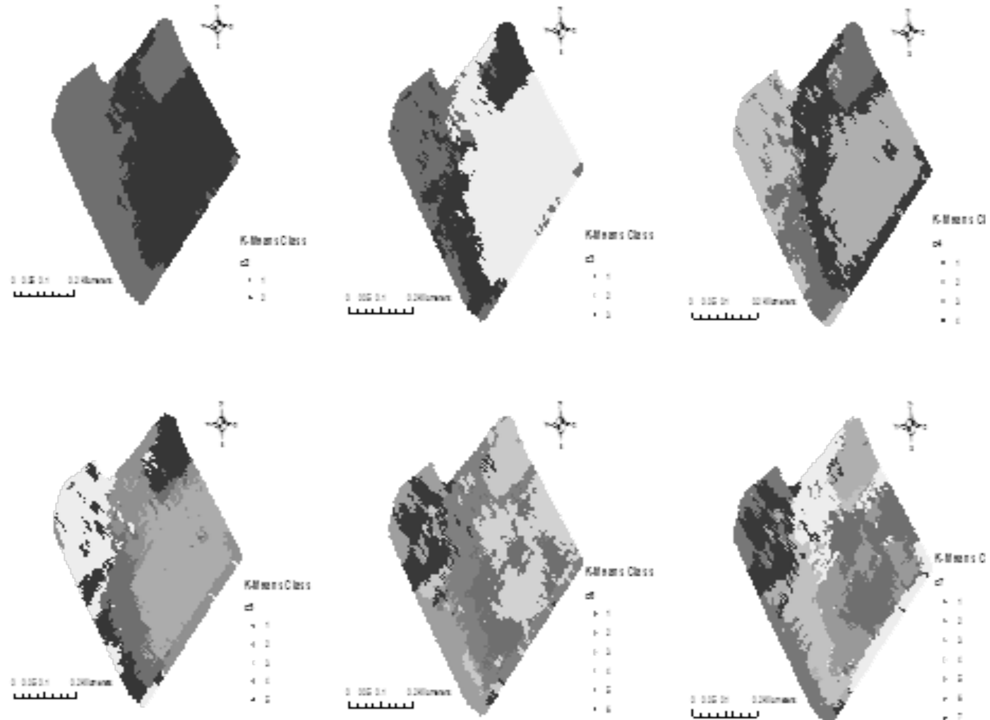

Fig. 1. The clusters formed by the k-means algorithm for the Horn's End dataset (in year (2013) data. The basic clusters are two while the left-side cluster is split in two further resulting in 3 clusters. The data presented here, is for clustering soil properties, NDVI and historic yield.

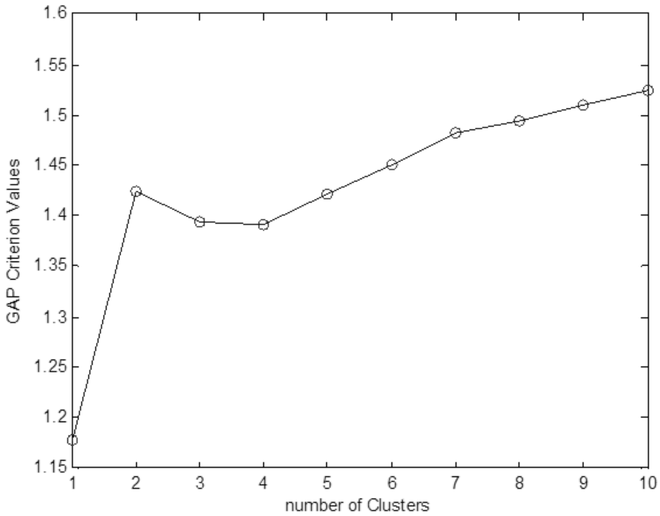

Fig. 2. GAP values for Horn's End

3.2 Data Fusion by Clustering with Self-Organizing Maps

The delineation of MZ by utilizing self-organizing maps (SOM) was achieved by using Matlab (Mathworks, Natick, MA, USA). The U-matrix was developed first before delineating MZ by applying the K-means algorithm on the U-matrix, resulting in MZ (Recknagel et al., 2006). The U-matrix represents the matrix of distances separating neighbors in the grid of SOM. The effectiveness of the U-matrix lays in its ability to visualize the neurons density in the data space by visual inspection of the distances between the clusters that neurons make in the weight space. In order to create maps of MZ, the sample data were supposed to belong to the group of neurons that are activated when these data are presented to SOM. The cluster formation seems to be clearer due to the fact that the SOM forms Voronoi polygons grouping similar vectors. Moreover, it gives a better view of the data microstructure letting the k-means to deal with higher level correlations of the data that is related to persistent phenomena which affect the data behavior. At this point, the clusters can be analyzed by U-matrix and dendrograms, as is shown in Figure 3.

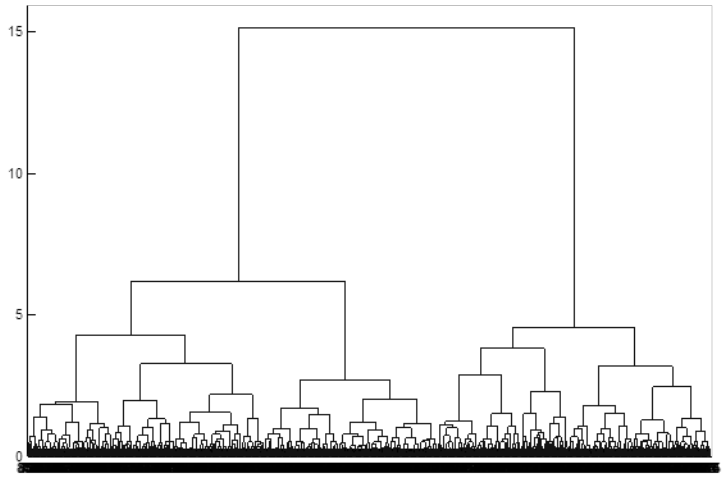

Fig. 3. The structure of the SOM clusters is shown in the dendrogram where for Horn's End – 2013 two major clusters are shown.

The k-means algorithm which is applied on top of the SOM clusters (Fig.4) demonstrates smoother interpolation of results as compared to the corresponding results produced with k-means clustering only which depend on the amount of Voronoi regions corresponding to the SOM neurons, forming the centroids of these regions. For example a 3x3 SOM with 9 Voronoi regions (polygons) results in the MZ maps shown in Fig.4.

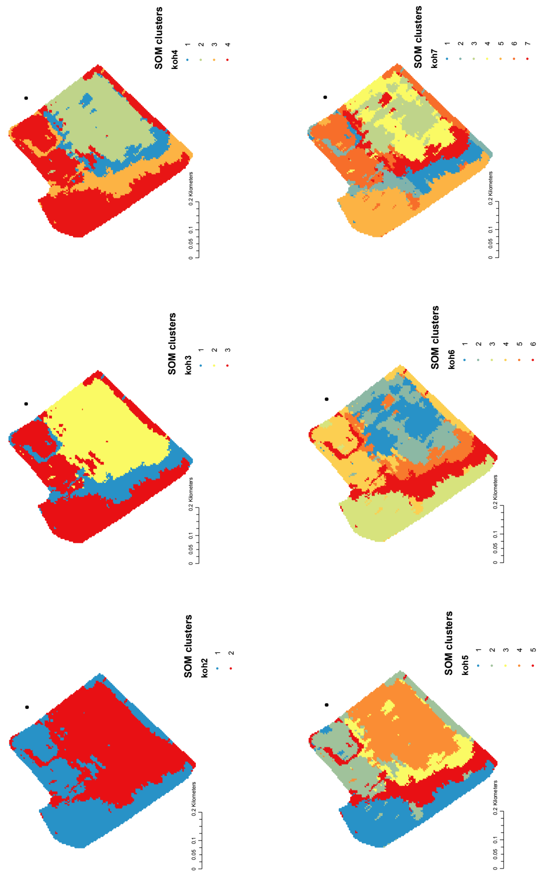

Fig. 4. The management zone maps produced by the combined self-rganizing map (SOM) with k-means algorithm for different number of clusters between 2 & 7. The basic clusters are two while the left-side of the field (blue cluster) is persistent during all subsequent segmentation indicating a serious anomaly in the data generation of physical phenomenon (probably also indicating a yield failure). This failure was due to water logging in this part of the field, where yield data were always low, although the soil fertility is high

In order to examine the goodness of separation between clusters resulting from the hybrid SOM and k-means clustering, the normalized mean plots for different variables can be examined. As can be seen from Figures 5 and 6, the normalized means exhibit a consistent trend for clusters with low yield in 2013 in both k-means and hybrid K-means and SOM clustering. However, in the case of the hybrid clustering, the normalized means of the soil parameters are well separated for all three clusters while in the case of k-means the topology of the means is distorted. This confirms the superiority of hybrid clustering regarding the separation between different classes compared to the corresponding K-means clustering.

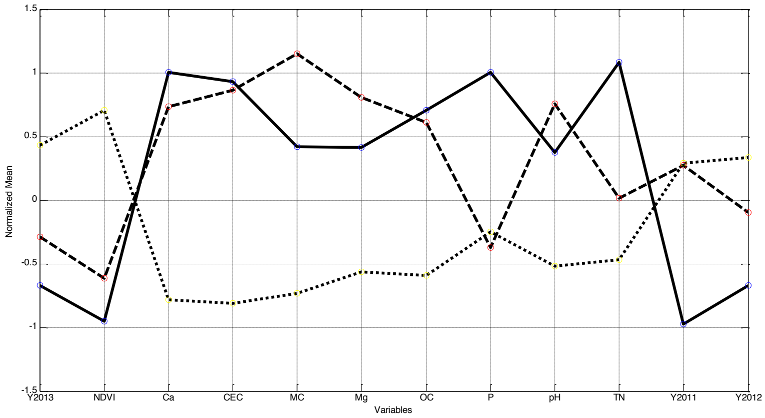

Fig. 5. Normalized means of K-mean Clusters

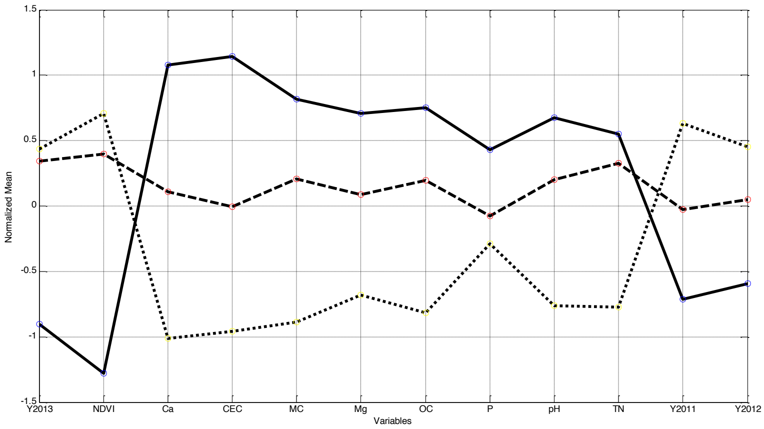

Fig. 6. Normalized Means of hybridSOM clusters (K-Means performed on SOM grid of neurons)

It is evident from the normalized means of the hybrid SOM clusters in Figure 6 that the low yield corresponds to high values of soil parameters. This can be explained from water logging problems in the corresponding areas of the field (left side of the field in Figure 4). The other two clusters demonstrate the inverse behavior where consistently lower values of soil parameter mean value relate to higher yields in 2013. This explains that although the soil fertility is high, the water logging problem prevents obtaining a good yield, whereas a lower level of soil fertility could result in a better yield when the soil is well-drained. A similar behavior can be observed concerning the NDVI, which seems to be highly correlated with the yield in all three clusters. The behavior of the yield is also consistent with yields of 2011and 2012.

4 Discussion

The cluster centers of the hybrid SOM and k-means algorithm show better separation of clusters when compared with the standard k- means algorithm. The cluster formation is clearer since the SOM forms Voronoi polygons grouping similar vectors and thus obtains a better view of the microstructure of the data allowing the k-means to deal with higher level correlations of the data related to persistent phenomena affecting the behavior of the data.

5 Conclusions

In this paper, the combination of common statistical approaches with Self Organizing Clustering for delineating MZ is presented. By this way, the management of the field related to the application of inputs is becoming more accurate since the relations of the soil and crop parameters are indicated in a more precise way. The soil parameters have been predicted based on proximal soil sensing utilizing high resolution spectral measurements and satellite based NDVI sensing. The obtained data layers have been fused and the point vectors have been subjected to clustering. The k-means algorithm is compared with the Self Organizing Map for delineating MZ. Further, a hybrid SOM algorithm is presented which forms clusters in combination with k-means. The hybrid SOM algorithm and k-means are compared in terms of cluster separation and MZ formation based on data fusion of Normalized Difference Vegetation Index (NDVI) and soil parameters. The cluster centers of the hybrid SOM and k-means algorithm show better separation of clusters when compared with the standard k- means algorithm.

Acknowledgements. The presented research was carried out in the framework of project FARMFUSE of ICT AGRI 2 ERANET.

References

1. Adamchuk, V. I., Hummel, J. W., Morgan, M. T. and Upadhyaya, S. K. 2004. On-the-go soil sensors for precision agriculture, In: Computers and Electronics in Agriculture, vol. 44, no. 1, pp. 71-91.
2. Boginski,V., Commander, C., Pardalos P.M., and Ye, Y. 2012. "Sensors: Theory, Algorithms, and Applications," Springer.
3. Chang, C. W., Laird, D. A., Mausbach, M. J., and Hurburgh, C. R. 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65, 480–490.
4. Christy, C. D. 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy, In: Computers and Electronics in Agriculture, vol. 61, no. 1, pp. 10-19.
5. 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. Am. Soc. Agric. Eng. 44 (1), 155–166.
6. Hartigan, J. A. and M. A. Wong 1979. Algorithm AS 136: A k-means clustering algorithm. In: Applied Statistics 28.1, pp. 100{108.
7. Kuang, B., Mahmood, H. S., Quraishi, M. Z., Hoogmoed, W. B., Mouazen, A. M. and van Henten, E. J. 2012. Chapter four - sensing soil properties in the laboratory, in situ, and on-line: a review, In: Donald Sparks, editor: Advances in Agronomy, vol. 114, Academic Press, 30 Corporate Drive, Burlington, MA 01803, USA , pp. 155-223.
8. Kuang, B. & Mouazen, A. M. 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms, In: European Journal of Soil Science, vol. 62, no. 4, pp. 629-636.
9. Mouazen, A. M., De Baerdemaeker, J. and Ramon, H. 2005. Towards development of on-line soil moisture content sensor using a fibre-type NIR spectrophotometer, In: Soil and Tillage Research, vol. 80, no. 1-2, pp. 171-183.
10. Recknagel, F., Talib, A., Van der Molen, D. 2006. Phytoplankton community dynamics of two adjacent Dutch lakes in response to seasons and eutrophication control unravelled by non-supervised artificial neural networks. Ecological Informatics, v. 1, n. 3, p. 277-285, ISSN 1574-9541.
11. Reyniers M (2003). Precision farming techniques to support grain crop production. PhD Thesis, Faculty of Applied BioSciences. Katholieke Universiteit Leuven, Belgium
12. Shibusawa, S., Anom, S. W. I., Sato, S., Sasao, A. and Hirako, S. 2001. Soil mapping using the real-time soil spectrophotometer, Proceedings of the 3rd European Conference on Precision Agriculture, (on CD-ROM), pp. 18.

13. Taylor J C; Wood G A; Earl R; Godwin R J (2003). Soil factors and their influence on within-field crop variability—part II: spatial analysis and determination of management zones. Biosystems Engineering, 84(4), 441–453

14. Tibshirani, R., Walther, G. and Hastie, T. 2001. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63: 411–423. doi: 10.1111/1467-9868.00293.

15. Yatsenko, V., Pardalos, P.M. and Kochubey S.M. 2003. Development of the Method and the Device for Remote Sensing of Vegetation, Remote Sensing for Agriculture, Ecosystems, and Hydrology IV (Edited by Manfred Owe, Guido D'Urso, and Leonidas Toulios), Proceedings of SPIE, Vol. 4879.