Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/092L08

On the Influence of Temporal Resolution on Automatic Delimitation using Clustering Algorithms

R. B. Arango[1], *I. Díaz*[1,*], *A. M. Campos*[1], *E. F. Combarro*[1] *and E. R. Canas*[2]

[1] Computer Science Department, University of Oviedo, Spain
[2] Technical Direction, Bodegas Terras Gauda, Spain

Received: 8 Jul. 2014, Revised: 9 Oct. 2014, Accepted: 10 Oct. 2014
Published online: 1 Apr. 2015

Abstract: One of the first steps in the application of precision agriculture techniques to a particular geographical zone is the task of land delimitation: determining which regions share similar soil properties and can (and should) be treated in a uniform way. In particular, automatic land delimitation is focused on providing delimitations from different data sources as those from satellite or sensors. In this work it is proposed to automatically delimit zones using remote sensed reflectivity and clustering algorithms. In addition, it is studied how temporal resolution affects this delimitation. In order to obtain this zoning, two different clustering paradigms are applied to data collected from Terras Gauda vineyard, a well known Spanish producer of Albariño wine. The results are promising in the sense that the clusters obtained are consistent with the current land organization and show that the lower temporal resolution, the more compact the clusters.

Keywords: Precision Agriculture, Land Delimitation, Clustering, Satellite Data

1 Introduction

Precision Agriculture (PA) or Precision Farming takes advantage of Information and Communications Technology (ICT) in order to provide valuable information and services to farmers. The term "precision" implies that these services can be customized to the needs of an area or specific farm plot and their characteristics, such as the spatial variability of the land and the different features at topographic or geological level. For instance, for the application of fertilizers, knowing the soil nutrients concentration and the in-field variability will allow to choose the right amount of fertilizer [1]. In a similar way, pest control can be more efficient by applying pesticides in a localized manner, where necessary, in contrast to its widespread application on the entire crop [2]. Among the tasks covered by PA, the identification of homogeneous zones of crop land areas is a key factor [3]. These management zones (MZ) [4] address spatial variability of crops grouping areas that share similar soil properties in order to apply specific farming practices to each MZ.

The knowledge of the farmer about the crops and soil could be a starting point in zones determination. However, other approaches provide methods for a systematic MZ identification such as the classification of apparent soil electrical conductivity [5] or the analysis of yield maps [6].

Sensors and onboard satellite instruments related to environmental and Earth observation measure electromagnetic radiation emitted and reflected by the observed objects. Based on these radiometric data it is possible to obtain valuable variables and indicators from the agriculture perspective [7] such as moisture and soil temperature, the vegetation index or even the kind of vegetation and its health. These remotely sensed data may be used for the estimation of soil properties and the recognition of spatial patterns [8]. However, non-commercial satellite data products with both high spatial and high temporal resolution are not available.

PA uses different measuring techniques such as small-size unmanned air vehicles equipped with spectrometers, luxometers or multi-spectral, modified land vehicles or sensor networks in the field. Using data from wireless sensor networks in greenhouse crops it is possible to develop disease early-warning systems [9] based on models considering leaf moisture, temperature and time factors. Advances in wireless sensors that use

* Corresponding author e-mail: sirene@uniovi.es

technologies such as RFID and Very Long Range Identification Tag [10] can help mitigate some drawbacks of mobile sensor networks [11] as their high cost or low autonomy.

However, the deployment of these measurement systems is not essential for PA. The use of satellite imagery and aerial photographs for measuring the radiation emitted and reflected by the fields in several areas of the electromagnetic spectrum allows to observe multiple variables which affect crops [12] as well.

Among the smart services that PA can provide, automatic land delimitation is one of the most challenging tasks. In fact, clustering and automatic delimitation of agroecozones (geographic zones that share similar ecological and environmental features) may be relevant to determine crops potentially more suitable for each zone.

This paper addresses automatic delimitation of the land, selecting algorithms to cluster land points characterised by the satellite reflectance data in order to evaluate how temporal resolution affects the clustering. The proposed method is tested by a case study for the grape vine crops of Terras Gauda, a producer from Galicia (Spain). Section 2 provides an overview of the precision agriculture approach. Section 3 explains the satellite data products used in the clustering process. Section 4 describes the approaches used to obtain the clusterings. Section 5 exposes results of the application of clustering algorithms. Finally, Section 6 shows the conclusions and presents some ideas for future work.

2 Automatic Land Delimitation

As it was stated before, in this work the problem of automatic land zoning in the viticulture environment is addressed. Automatic zoning has been largely studied, specially using sensor data as a tool to detect homogeneous zones. For example, Le Ber in [13] used satellite imagery to build an expert system prototype which is able to recognize the land-use and to calculate its geometrical properties such as perimeter, surface area and form. Ortega and Santibaez in [14] systematically delimited crop management zones relying on six soil chemical properties related to fertility and performing clustering analysis, principal component analysis and using the coefficient of variance concluding that no significant differences were found on the results obtained with those three methods.

Simbahan y Dobermann [15] proposed supervised spatial classification algorithms to identify MZ using as parameters the number of clusters, its minimum size and the weights of the categorical variables. The method was tested with different datasets including soil maps, digital elevation models, apparent soil electrical conductivity and soil surface.

Other approaches [16] applied the k-means clustering algorithm to remote sensing data obtained from the MODIS-based greenness index [17] and also to the seasonal leaf area index [18]. The authors developed a parallel algorithm implementation and compared computational time according with the number of target clusters and the number of parallel processors. As an example, 1,000 agroecozones were delimited in 700 seconds using 2,048 processors with their clustering implementation.

In the same line, Kumar et al. [19] applied k-means to Identify homogeneous zones for a cotton field, considering two datasets: the first one with two estimators of yield and the second one considering geo-referenced field properties such as topographical characteristics and treatments applied to the field.

A fuzzy clustering approach was proposed by Liu and Samal [20] considering biophysical features such as annual moisture deficit/surplus and mean annual precipitation. The authors tested the same dataset with k-means and fuzzy algorithms concluding that A fuzzy approach generates more accurate delineations.

3 Input Data

Valuable data related to PA such as vegetation indices, land surface temperature or surface reflectance are collected by MODIS (Moderate Resolution Imaging Spectroradiometer) [21]. Around 70 data products are provided by this instrument operated from TERRA and AQUA satellites [18]. Table 1 shows a sample of the MODIS data products related to precision agriculture. These data products are publicly available by HTTP, FTP and at the NASA Land Processes Distributed Active Archive Center [17]. Usually the data will consist of images on JPG format, XML files and data in a hierarchical format (HDF).

In order to generate the data sets for the MZ identification process, daily surface reflectance data products were considered; specifically, MOD09GQ at 250m of spatial resolution. It includes data about surface reflectance for spectral bands 1 and 2 and other variables to measure the quality of the observations and their coverage [22].

Table 1: A sample of MODIS data products related to Precision Agriculture

Name	Data Product	Res. (m)	Frequency
MYD09GA	Surface Reflectance Bands 1-7	500 m	Daily
MOD09GQ	Surface Reflectance Bands 1-2	250 m	Daily
MOD11A1	Land Surface Temp. and Emissivity	1000 m	Daily
MOD13Q1	Vegetation Indices	250 m	16 days
MOD15A2	Leaf Area Index - FPAR	1000 m	8 days
MOD14A1	Thermal Anomalies and Fire	1000 m	Daily
MOD44B	Vegetation Continuous Fields	250 m	Annual

Table 2: Columns of the dataset for the MOD09GQ data product. Spatial resolution: 250 m. Temporal resolution: daily. CRS: UTM 29

Column	Description
x	Coordinate x of the data point in the UTM 29 CRS
y	Coordinate y of the data point in the UTM 29 CRS
date	Year + Day number of the year in the format YYYYddd
refl_b01	Reflectivity values from MOD09GQ band 1
refl_b02	Reflectivity values from MOD09GQ band 2
num_observations	The number of observations
QC_250m	A byte about the quality of the measure
NDVI	Normalized Difference Vegetation Index
NDVI_scaled	Scaled NDVI

NDVI indicator is calculated as the relation between the difference of the values of both Red (Red) and NIR (near infrared) channels and its sum [23].

$$NDVI = (NIR - Red)/(NIR + Red)$$

Negative values correspond to water, clouds or snow since their reflectance in the visible spectrum is greater than the corresponding in near infrared, whilst soil and rocks have values near zero. Ranges between 0.1 and 0.6 are indicators of vegetation. Values above 0.6 correspond to dense vegetation canopy.

4 Clustering Algorithm

As it was described in Section 3, the data used in this approach include information about surface reflectance for spectral bands 1 and 2 with a spatial resolution of 250m obtained from MODIS [22] and the NVDI and NVDI-scaled indices. To obtain the automatic delimitation of the vineyard zones two different approaches are used. The first one is based on a partition clustering algorithm called Partitioning Around Medoids (PAM). The second one is based on hierarchical clustering paradigm. Next subsections briefly describe both approaches.

4.1 PAM clustering algorithm

The main characteristics of PAM [24] are the following:

– It is a partitioning algorithm. Thus, it breaks the input data up into groups until some stability condition is reached.
– The number of groups is defined in advance.
– PAM stands for Partition Around Medoids. It tries to find a set of objects called medoids that are centrally located in clusters.
– PAM is an algorithm more robust than k-means because it minimizes a sum of dissimilarities instead of a sum of squared euclidean distances.

The main difference between this algorithm and the classical k-means method is that PAM uses medoids as centers of the clusters and these medoids are selected among the objects to be clustered. The algorithm involves three steps:

1. **Initialization**: Select, at random, the k medoids from the data points
2. **Assignment**: For each point, locate the closest medoid and assign it to the corresponding cluster
3. **Update**: For each cluster, compute the new medoid from the points assigned to the cluster. The new medoid will be the point that minimizes the dissimilarity to the rest of the elements in the cluster.

Steps 2 and 3 are repeated until the clusters are no longer modified.

The bottleneck of clustering algorithms is to properly select the best clustering distribution. In fact, the evaluation of clustering structures is the most difficult task in clustering algorithms. A large number of ways of evaluating the goodness of a clustering algorithm have been proposed in the literature. In this case, since we have no reference to external information, the method selected to validate the clustering was the Silhouette coefficient [25]. It is based on the comparison of cluster tightness and separation. This Silhouette shows which objects lie well within their cluster, and which ones are merely somewhere in between clusters. The average silhouette width provides an evaluation of the clustering validity, and can be used to select an appropriate number of clusters.

Therefore, to select the optimum k according to the Silhouette coefficient, we follow the procedure described below (suppose that the number of points to cluster is n and that K^* is the maximum number of clusters, which is equal to or less than n):

for $j = 1, K^*$ **do**

for $i = 1, n$ **do**

$a(i) = \frac{\sum_{j \in C_i}(d(i,j))}{\#C_i}$, C_i cluster of element i

$b(i) = min_C d(i,C), \forall C$

$s(i) = \frac{b(i)-a(i)}{max\{(b(i),a(i)\}}$,

end for

$s^j_{avg} = \frac{\sum_{i=1}^{n} s(i)}{n}$

end for

$k = argmax\{s^j_{avg}\}$

Once the number of clusters is computed, the cluster assignment is retrieved, providing the land delimitation.

4.2 Hierarchical clustering

Another interesting approach to clustering is based on hierarchical methods. Cluster analysis is often used to bring similar individuals into groups. In hierarchical clustering, individuals are successively integrated based on the dissimilarity matrix computed from the data, to obtain a dendrogram which contains inclusive clusters. Among all the hierarchical algorithms, *pvclust* algorithm was selected because it is a combination of the standard hierarchical algorithm *hclust* and bootstrap resampling so the groups it obtains have higher confidence [26].

Hierarchical algorithm *hclust* uses a set of dissimilarities for the objects being clustered. Initially, each object is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. At each stage distances between clusters are recomputed to the particular clustering method being used.

There are many different clustering methods to apply. Ward's minimum variance method aims at finding compact, spherical clusters. The complete linkage method finds similar clusters. The single linkage method is closely related to the minimal spanning tree. According to the semantic of the problem being solved here, the most adequate method is the Ward one [26]. The output of this method is a dendrogram, showing not only the groups but also the strength of the connection between them.

However, it is not clear how strongly a cluster is supported by the data. To check the certainty of the existence of a cluster, p-values are computed. If the p-value is less than a certain threshold (usually a number close to 1), the cluster is rejected. The p-values are computed with multiscale bootstrap resampling.

5 Experiments

5.1 Clustering with PAM

The purpose of this paper is to automatically delimit the Terras Gauda vineyard. The Terras Gauda vineyard is divided into three separated parcels (hereinafter called p_1, p_2, p_3). According to the spatial resolution provided by the MODIS satellite, that is 250m., p_1 and p_2 are represented using 30 points and p_3 by 16 points. Each point is characterised by 4 variables per day (surface reflectance for spectral bands 1 and 2, NVDI and NVDI-scaled indices). As data were extracted for 90 days, each point x is represented by 360 values as follows

$$x = (b1_{day_1}, b2_{day_1}, NVDI_{day_1}, NVDIscaled_{day_1}, b1_{day_2}, b2_{day_2}, NVDI_{day_2}, NVDIscaled_{day_2}, \ldots, b1_{day_{90}}, b2_{day_{90}}, NVDI_{day_{90}}, NVDIscaled_{day_{90}})$$

Following the procedure described in Section 4, we have clustered the three parcels using the Manhattan distance as dissimilarity metric. We have tested the performance of the values obtained from the MODIS satellite in automatic land delimitation and also the effect of different temporal resolutions. To test the temporal resolution we consider all the attributes every day, every 2 days, ... until every 10 days.

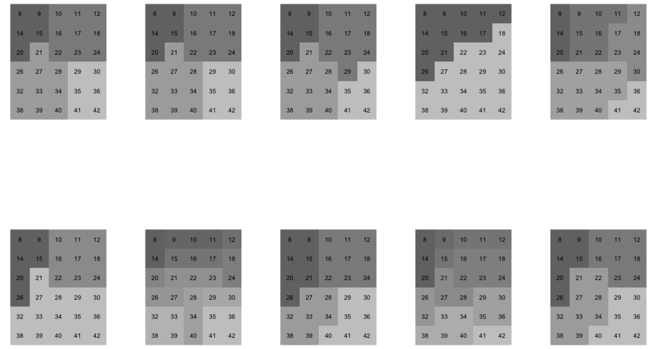

Fig. 1: Land delimitation for parcel p_1.

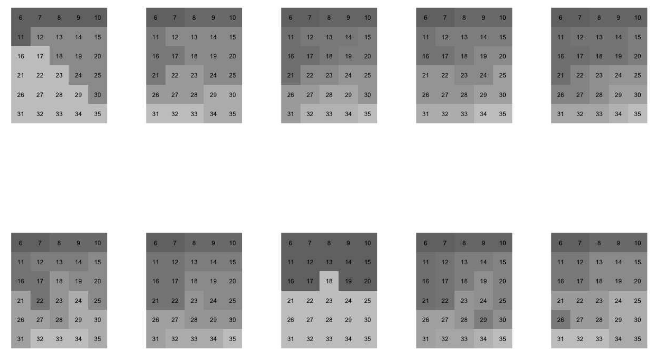

Fig. 2: Land delimitation for parcel p_2.

To test the effect of the four attributes, the clustering procedure was performed considering each possible combination of the four attributes obtained per day. However, since the results are similar, only the clusterings obtained when the four attributes afore defined are taken together will be considered in what follows.

Figures 1 to 3 show the land delimitation for parcels p_1, p_2 and p_3, respectively. The structure of each figure is the following: It contains 10 squares divided into smaller squares according to the the spatial resolution provided by MODIS. Considering each figure as a matrix with 2 rows and 5 columns, the square at position $[1, j]$ contains the clusters obtained when temporal resolution is j. The square at position $[2, j]$ contains the clusters obtained when temporal resolution is $j+5$. Therefore the topleft square represents clustering results for daily resolution and the bottomright square represents clustering results for a ten days resolution.

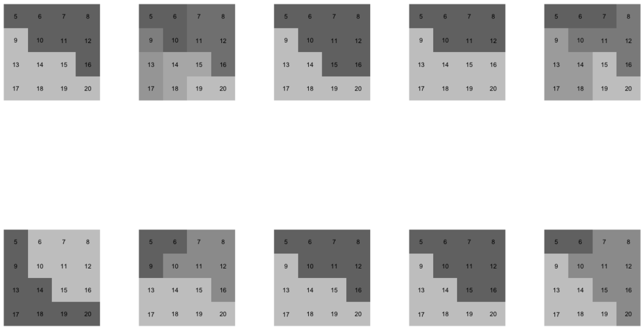

Fig. 3: Land delimitation for parcel p_3.

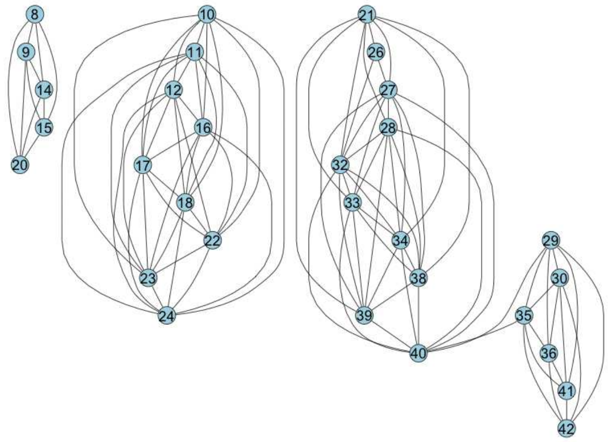

Fig. 5: Connected components for parcel p_1 with $h = 5$

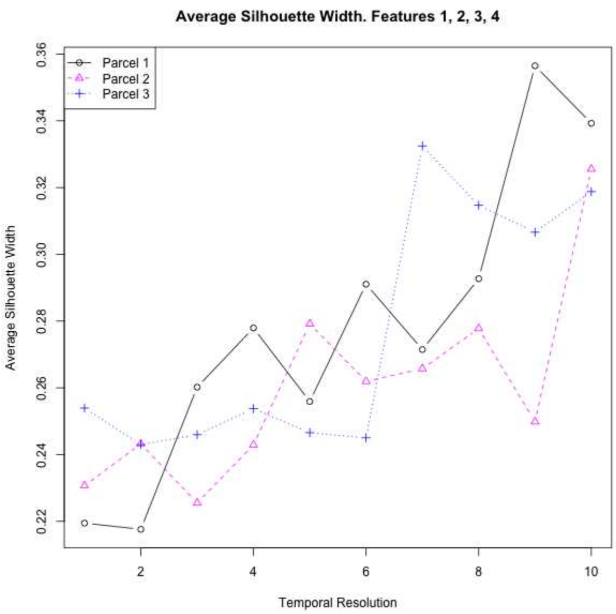

Fig. 4: Silhouette coefficient obtained by the clustering algorithm

The question now is how to decide which land delimitation is the best. As it was stated in Section 4, the criterion used is the Silhouette coefficient. X-axis of Figure 4 represents the temporal resolution (from 1 day to 10 days) while Y-axis represents the value of the Silhouette coefficient. As it can be seen in Figure 4 the lower the temporal resolution, the higher the Silhouette coefficient. Therefore, the main result we can extract is that there is no need to include daily information in order to keep the performance of the clusters. In fact, the clusters are more compact when the temporal resolution is lower.

5.1.1 Clustering aggregation

In order to obtain a better land delimitation from the ones obtained in the previous subsection with different temporal resolutions, we have explored the possibility of aggregating the different clusterings. To this extent, we have considered a series of binary relations R_h defined as follows: two points a_1 and a_2 are related by R_h (in symbols, $a_1 R_h a_2$) if and only if they appear together in at least h of the clusterings considered in the previous subsection.

Thus, we he have ten different binary relations $R_1, R_2, \ldots, R_{10}$ each one contained in the previous one (i.e., if $i < j$ then $a_1 R_j a_2$ implies $a_1 R_i a_2$). We consider, then, the graph representation of these binary relations and calculate their connected components. Each of these connected components contains points that are usually clustered together (with a higher probability the higher the value of h is) and, in consequence, may be considered as a new, aggregated clustering.

Figures 5 to 7 show some of these graphs for parcels p_1 to p_3. In the case of the two first parcels we have set $h = 5$ while h is 6 for the third case.

As can be seen, in these cases we obtain 3 areas for parcels p_1 and p_2 and 2 for p_3 which may be considered as *consensus* areas between all the clusterings (cf. Figures 1 to 3).

5.2 Clustering with pvClust

In this section it is studied the problem of land delimitation using a hierachical clustering method. As in the previous case, we just show the results obtained when using the four input variables, testing the influence of temporal resolution.

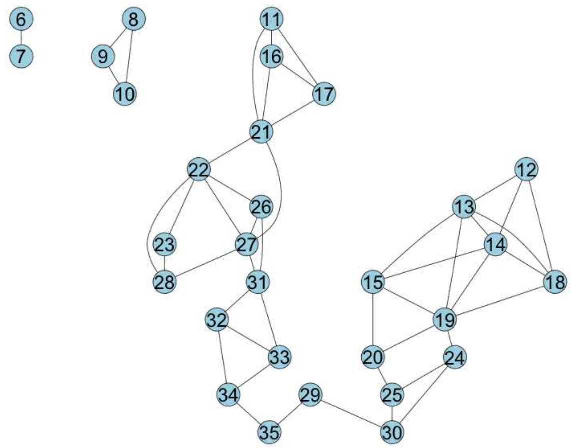

Fig. 6: Connected components for parcel p_2 with $h = 5$

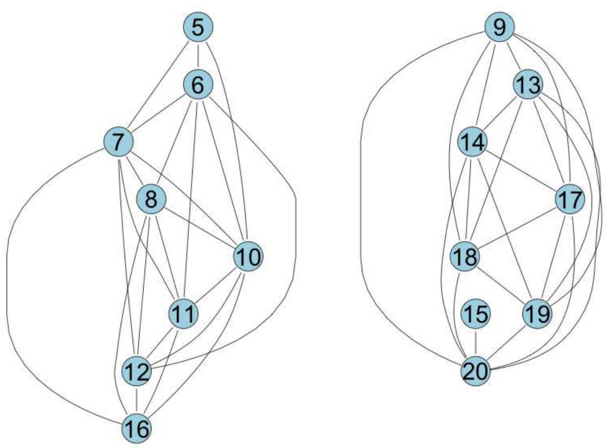

Fig. 7: Connected components for parcel p_3 with $h = 6$

As it was detailed in Section 4.2, this method provides us with a parameter p to check the certainty of the existence of a cluster. Therefore, this parameter is used to select the best temporal resolution. Figures 8 to 10 represent p-values against standard error for the best temporal resolution for parcels p_1, p_2 and p_3 respectively.

Note that the higher the number of p-values close to 1, the better. In addition, the lower the standard error, the better. According to these premises, the best resolutions are 9 for parcel p_1, 7 for parcel p_2 and 10 for parcel p_3. Figures 11 to 13 show the dendrograms associated to the selected resolutions obtained by *pvclust* for the three parcels. Groups with a certainty above 90% are highlighted. It is important to note that in general the lower the temporal resolution, the higher the p-values. This fact allows to conclude that when temporal resolution is lower, the performance of the clusterings is at least the same that the one obtained with all data. Therefore, there is no need to use daily data.

With regard to the groups themself, note that for parcel p_1 the system is not able to group with a high confidence only four points (9, 12, 20 and 21) , five for parcel p_2 and two for parcel p_3.

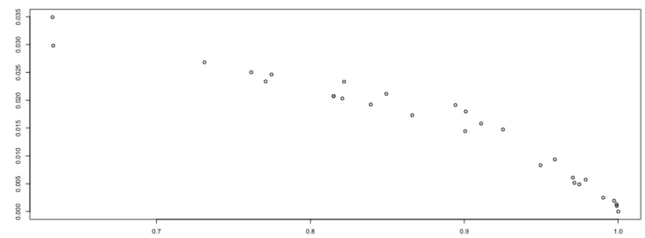

Fig. 8: p-value against standard error for parcel p_1 and data obtained each 9 days

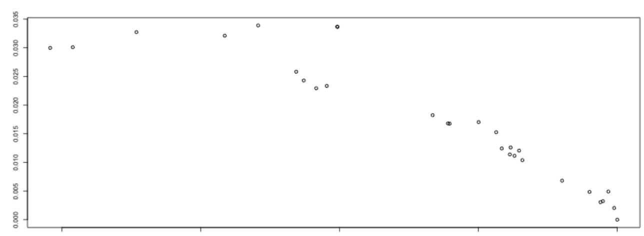

Fig. 9: p-value against standard error for parcel p_2 and data obtained each 7 days

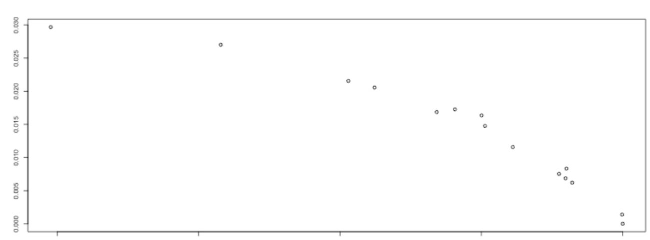

Fig. 10: p-value against standard error for parcel p_3 and data obtained each 10 days

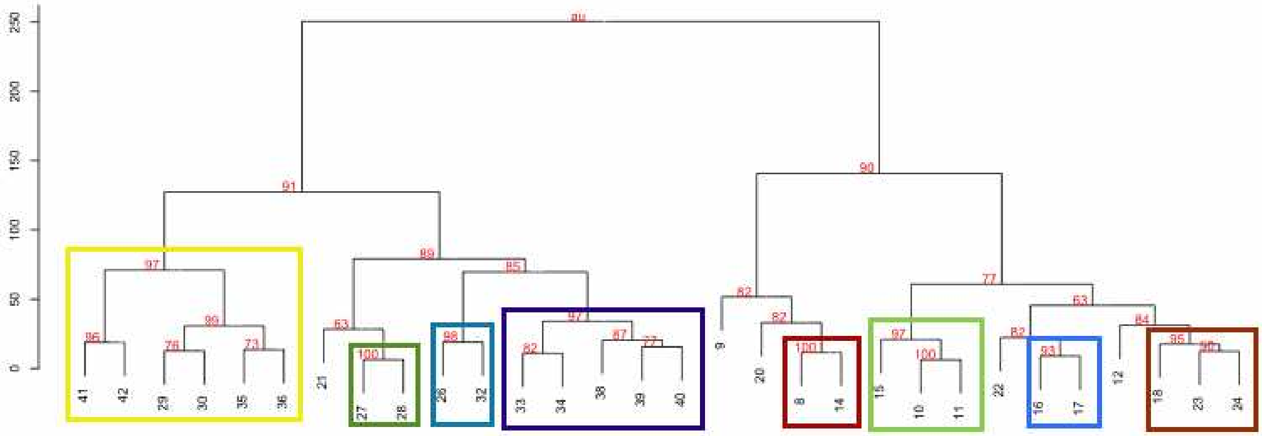

Fig. 11: Dendrogram representing parcel p_1 clustering

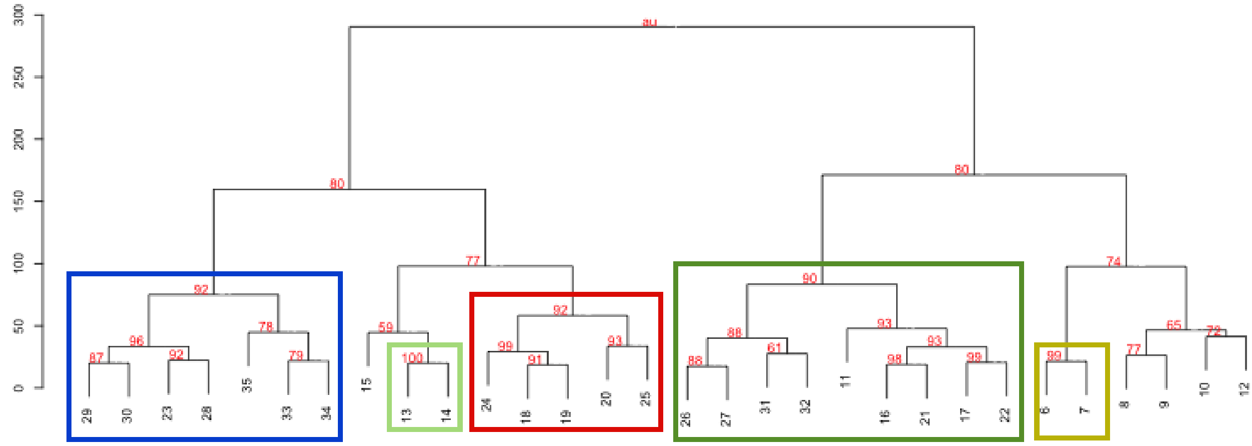

Fig. 12: Dendrogram representing parcel p_2 clustering

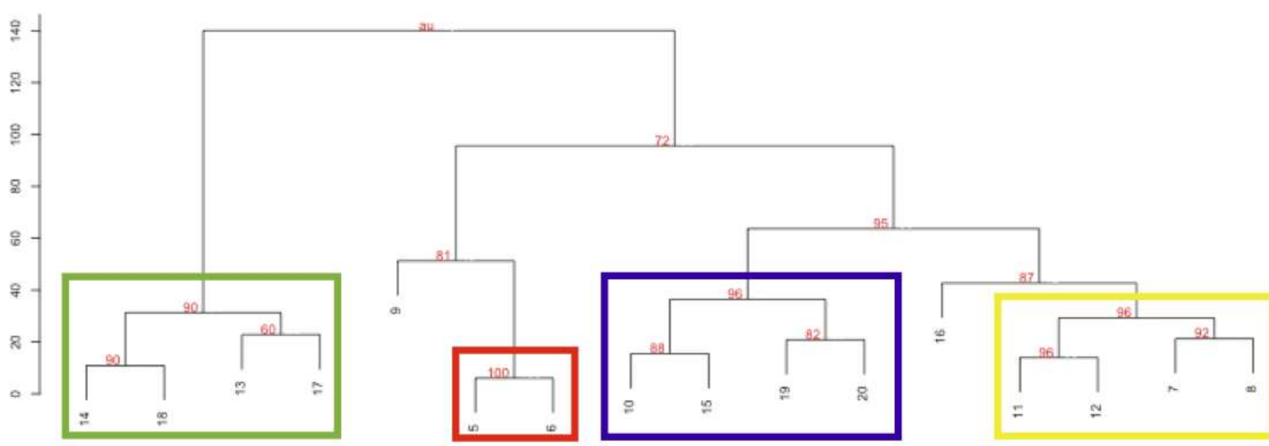

Fig. 13: Dendrogram representing parcel p_3 clustering

6 Conclusions and Future Work

This paper presents a first attempt to automatic delimitation of Terras Gauda vineyard. Terras Gauda is a well known Spanish producer of Albariño wine. The results are promising in the sense that it was checked that a lower temporal resolution does not affect the land delimitation. Actually, the results show that the lower the resolution, the more compact the clusters. This work opens many interesting new problems. Among them land delimitation using data retrieved from Landsat 8, whose spatial resolution is much higher but the temporal one is lower. Another future line is to study fusion techniques in order to improve the clustering.

Acknowledgements

This work has been partially supported under MEC and FEDER Grant TEC2012-38142-C04-04, UNOV-13-EMERG- GIJON-10 grant from University of Oviedo and Pilot B no.621074 from European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under grant agreement no. 621074.

References

[1] H. Yu, D. Liu, G. Chen, B. Wan, S. Wang, B. Yang, A neural network ensemble method for precision fertilization modeling, Mathematical and Computer Modelling 51 (11) (2010) 1375–1382.

[2] C. Fernández-Quintanilla, J. Dorado, C. San Martín, J. Conesa-Muñoz, A. Ribeiro, A five-step approach for planning a robotic site-specific weed management program for winter wheat, Robotics and Associated High-Technologies and Equipment For Agriculture, 2011.

[3] A. R. Schepers, J. F. Shanahan, M. A. Liebig, J. S. Schepers, S. H. Johnson, A. Luchiari, Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years, Agronomy Journal 96 (1) (2004) 195–203.

[4] R. Ferguson, R. Lark, G. Slater, Approaches to management zone definition for use of nitrification inhibitors, Soil Science Society of America Journal 67 (3) (2003) 937–947.

[5] C. K. Johnson, D. A. Mortensen, B. J. Wienhold, J. F. Shanahan, J. W. Doran, Site-specific management zones based on soil electrical conductivity in a semiarid cropping system, Agronomy Journal 95 (2) (2003) 303–315.

[6] S. Blackmore, R. J. Godwin, S. Fountas, The analysis of spatial and temporal trends in yield map data over six years, Biosystems engineering 84 (4) (2003) 455–466.

[7] S. Ormeño Villajos, A. Arozarena Villar, M. Martínez Peña, M. Palomo Arroyo, G. Villa Alcázar, J. Peces Morera, L. Pérez García, Los satélites de media y baja resolución espacial como fuente de datos para la obtención de indicadores ambientales, in: IX Congreso Nacional de Medio Ambiente, Madrid (In Spanish), 2008.

[8] A. Bhatti, D. Mulla, B. Frazier, Estimation of soil properties and wheat yields on complex eroded hills using geostatistics and thematic mapper images, Remote Sensing of Environment 37 (3) (1991) 181–191.

[9] M. Neto, F. Baptista, L. Navas, G. Ruiz, A business intelligence approach to support a greenhouse tomato crop grey mould disease early warning system, in: T. Mildorf, K. C. jr. (Eds.), ICT for Agriculture, Rural Development and Environment, Czech Centre for Science and Society, 2012, pp. 175–184.

[10] Z. Krivanek, K. Charvat, J. Jezek, M. Musil, VLIT NODE sensor technology and prefarm, AGRIS on-line Papers in Economics and Informatics 2.

[11] P. Kubíček, V. Lukas, J. Kozel, Selected issues of wireless sensor networks geovisualization in agriculture, in: ICT for Agriculture, Rural Development and Environment, 2012, pp. 249–263.

[12] W. Bingfang, M. Jihua, Z. FeiFei, D. Xin, Z. Miao, C. Xueyang, Applying remote sensing in precision farming-a case study in Yucheng, in: World Automation Congress, 2010, pp. 1–6.

[13] F. Le Ber, A prototype model-based expert system for agricultural landscape analysis, AI Applications-Natural Resources, Agriculture, and Environmental Science 9 (2) (1995) 91–101.

[14] R. A. Ortega, O. A. Santibáñez, Determination of management zones in corn (Zea mays L.) based on soil fertility, Computers and Electronics in agriculture 58 (1) (2007) 49–59.

[15] G. C. Simbahan, A. Dobermann, An algorithm for spatially constrained classification of categorical and continuous soil properties, Geoderma 136 (3) (2006) 504–523.

[16] J. Kumar, R. T. Mills, F. M. Hoffman, W. W. Hargrove, Parallel k-means clustering for quantitative ecoregion delineation using large data sets, Procedia Computer Science 4 (2011) 1602–1611.

[17] NASA land processes distributed active archive center (LP DAAC). ASTER L1B. USGS/Earth resources observation and science (EROS) center, Sioux Falls, South Dakota. 2001.

[18] L. DAAC, MODIS products table [cited June, 2013].
URL http://lpdaac.usgs.gov/products/modis_products_table

[19] E. Schuster, S. Kumar, S. E. Sarma, J. Willers, G. Milliken, Infrastructure for data-driven agriculture: identifying management zones for cotton using statistical modeling and machine learning techniques, in: Emerging Technologies for a Smarter World (CEWIT), 2011 8th International Conference Expo on, 2011, pp. 1–6. doi:10.1109/CEWIT.2011.6163052.

[20] M. Liu, A. Samal, A fuzzy clustering approach to delineate agroecozones, Ecological modelling 149 (3) (2002) 215–228.

[21] L. P. D. A. A. Center, MODIS overview [cited October 2013].
URL https://lpdaac.usgs.gov/products/modis_overview

[22] MOD09GQ [cited May 2014].
URL http://goo.gl/K9D51L

[23] F. Kriegler, W. Malila, R. Nalepka, W. Richardson, Preprocessing transformations and their effects on multispectral recognition, in: Remote Sensing of Environment, VI, Vol. 1, 1969, p. 97.

[24] X. Li, K-Means and K-Medoid, in: L. Liu, M. T. Özsu (Eds.), Encyclopedia of Database Systems, Springer US, 2009, pp. 1588–1589.
URL http://dblp.uni-trier.de/db/reference/db/k.html#Li09f

[25] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (0) (1987) 53 – 65.

[26] F. Murtagh, P. Legendre, Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm.

Rodolfo de Benito is a PhD Candidate in Computer Science at University of Oviedo (Spain). He has a Masters degree in Computer Science at Polytechnic School of Engineering of Gijn, University of Oviedo and a degree on Computer Engineering at Open University of Catalua. His main research interests are in the areas of smart services, machine learning and predictive models for decision support in a big data context.

Irene Díaz is mathematician from the University of Oviedo since 1995. In 2001 she obtained her phd in Computer Science from University Carlos III of Madrid. Since 1999 she works for the Computer Science Department of the University of Oviedo, being senior lecturer since 2008. Her research interest are mainly focused on data mining, information retrieval and fuzzy logic. She has published more than 30 papers in journals indexed in SCI and more than 100 in conferences

Antonio Campos is the Director for Technology and Innovation of SERESCO since Nov. 2012 and part time professor of the University of Oviedo since 2005. He has a PhD in Informatics and a MSc. in Industrial Engineering by the University of Oviedo (Spain). Since 1996 Dr. Campos has participated in diverse international R&D initiatives at research or management levels, first as full-time associate professor of the University of Oviedo and from 2005 to 2011 as the Director for Research and innovation of CTIC Technological Center.

Elías F. Combarro received the PhD degree in mathematics from the University of Oviedo. He is an associate professor of computer science at the University of Oviedo, Spain. He has conducted research in computability theory, text categorization, automatic classification of algebraic structures, fuzzy measures and linear extensions of partially ordered sets.

Emilio Rodríguez Canas has a degree in Pharmacy from the University of Santiago de Compostela and a Masters degree in Enology and Viticulture from the ETSIA of Madrid. From the very beginning of his career he developed his work on both fields, enology and viticulture, assuming the Technical Direction on some winemaker companies and currently at Terras Gauda. Member of the committees of wine tasting and appellation of the protected designation of origin of Ras Baixas. Wine tasting Middle and Superior Grade Formative teacher at the School of Enology A Granxa at Ponteareas (Pontevedra, Spain). Award from the Galician Academy of Sciences 2009. Multiple research projects in collaboration with CSIC related with the selection of vine, the enological microbiology or the potential benefit effects for the wine health and their subproducts. These projects have allowed him to be listed as inventor on two join patents with CSIC and to license a third one, as well as to publish related papers in multiple journals.