Evaluation of time-series of satellite reflectance data for land delimitation using clustering algorithms

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

[1]Computer Science Department, University of Oviedo
[2]Terras Gauda

Motivation

- Identification of **homogeneous zones of crops based on remoted sensed reflectivity**
- Study how the temporal resolution affects to the delimitation of zones
- Apply it to Terras Gauda grapes crop.

Terras Gauda wineyard

PAM Clustering Algorithm

- 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 ***P*artition 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.
- Clustering performance metric: **Silhouette coefficient,** $s(i) = \frac{b(i)-a(i)}{max\{(b(i),a(i)\}}$, with $a(i) =$ average dissimilarity between i and all other points of the cluster to which i belongs and $b(i) = min_C d(i, C), \forall\, cluster\, C$
- Dissimilarity metric: **Manhattan distance**

Digital Data

- Collected by **MODIS** (Moderate Resolution Imaging Spectroradiometer) .
- 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.

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 Temperature 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

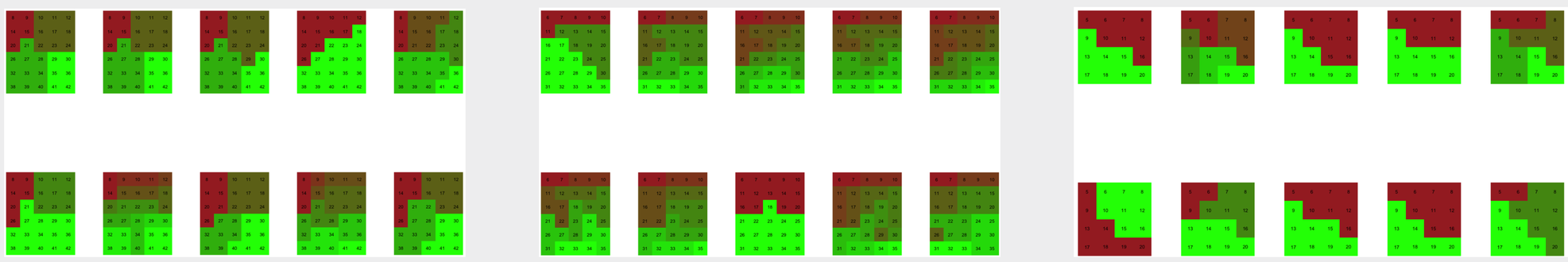
Table 1: **A sample of MODIS data products related to Precision Agriculture**

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 for this measure
QC_250m	A byte of information about the quality of the measure
NDVI	Normalized Difference Vegetation Index
NDVI_scaled	Normalized Difference Vegetation Index. Scaled [0..255] values

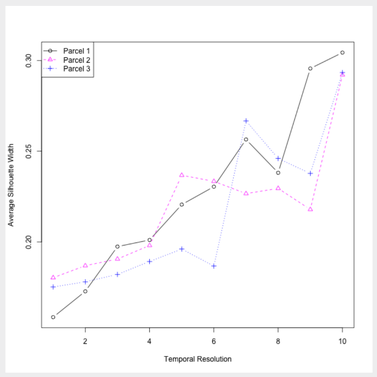
Table 2: **Columns of the dataset for the MOD09GQ data product. Spatial resolution: 250 m. Temporal resolution: daily. CRS: UTM 29**

Experiments

- **Terras Gauda** vineyard is divided into three separated parcels (p_1, p_2, p_3).
- p_1 and p_2 are represented using 30 points and p_3 by 16 points (according to MODIS resolution).
- Each point is characterised by 4 **variables per day** (surface reflectance for spectral bands 1 and 2, NVDI and NVDI-scaled indexes).
- As **data** were extracted for 90 **days**, each point x is represented by 360 **attribute** as follows $x = (b1_{day_1}, b2_{day_1}, NVDI_{day_1}, b1_{day_2}, b2_{day_2}, NVDI_{day_2}, \ldots, b1_{day_{90}}, b2_{day_{90}}, NVDI_{day_{90}})$
- Points are clustered using PAM algorithm using manhattan distance as similarity metric.
- The optimum number of clusters was computed according to the Silhouette coefficient.



From left to right figures respectively show the land delimitation for parcels p_1, p_2 an p_3. 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. Each cluster is represented by a different color.



Silhouette coefficient. X-axis represents the temporal resolution
Y-axis represents the value of the coefficient