Master in Space Applications for Emergency Early Warning and Response.

Crop classification in La Pampa by using multi-temporal and multi-sensor satellite data.

Soraya Violini ab, Claudia Notarnicola c

a Comisión Nacional de Actividades Espaciales (CONAE), Instituto Gulich, Córdoba, Argentina.
b Agenzia Spaziale Italiana (ASI), Italia.
c Institute for Applied Remote Sensing, European Academy (EURAC), Bolzano, Italia.

Key words: Landsat TM5, classification, methods, crop detection.

1. Introduction

Knowing the distribution and agricultural crop area is important for the planning of agricultural policies and the definition of rural land use bases, for researchers, to build models of ecosystem functioning and producers in making decisions (Bagnato et al., 2012). Traditionally, these data are derived from agricultural census, conducted in widely spaced periods of time, or to field studies, which are very expensive and do not cover large areas (Ding and Chen, 2010). However, in recent years, the advance of the remote sensing technology allows multitemporal studies of large areas from the use of satellite images.

The strategies to be followed for obtaining these maps are numerous and include the selection of the sensor type, the method of classification (supervised, unsupervised or mixed), the algorithm (parametric or nonparametric), stripe sets and / or combinations of them and the number of images (mono-temporal or multitemporal classifications), among others (Chuvieco, 2002, Lu and Weng, 2007).

In relation to the sensors, Ding and Chen (2010), demonstrated that the combination of images from optical and radar sensors provides more detailed and accurate information of the changes of vegetation cover and productivity and allows monitoring at different scales.

In terms of classification methods, numerous studies exploited the advantages and disadvantages of supervised and unsupervised methods (Larrañaga et al., 2010; Pajares et al., 2012; MacNairn et al., 2000, Wu et al., 2011, among others). The first ones are used when prior knowledge of the study area is available (Argañaraz, 2011) and is conducted in two phases: training and validation. By contrast, the unsupervised ones do not require any prior information about the scene and the analysis points to the discrimination of groups of objects that allow establishing different classes (Aiazzi et al., 2008).

The algorithm more used in supervised classification is the parametric Maximum Likelihood (Balenzano et al., 2011) that assumes data normality. However, the problems associated with non-compliance of this assumption have motivated the search for alternative nonparametric (Argañaraz, 2011). Among these, two algorithms have shown great performance in studies of land use in agricultural systems in satellite images: Hopfield Neural Network (HNN) verifies by calculating a coefficient of separability and a measure of homogeneity of the groups, taking into account two types of relationships between the pixel under consideration and its neighbors (Pajares et al., 2012) and Support Vector Machines (SVM). This method is capable of forming a decision about the domain of the training data with little or no knowledge of the domain of the data outside this boundary, with the description given of support vector data (Betancourt, 2005).

This work tries to do a comparison of the most used methods in supervised classification of land use (Maximum Likelihood) and new methods
used to optimize them (HNN and SVM), combining the use of images of optical and radar sensors for crop classification.

2. Data and study area

2.1. Study Area

The study area is located in the Province of La Pampa (Argentina), Capital department, in the vicinity of the locality of Anguil (Figure 1), between the coordinates 64°10'25.86"S and 63°40'34.63"O 36°18'26.57"S. The soils are of sandy texture, with wavy relief composed by hillocks and depressions that can still generate lagoons and saltpeter beds. The dominating soil type is a *Haplustol éntico*, family franc mixed gross, with a simple profile of the type A-AC-C-tosca, a deep arable layer, good provided with organic matter and well structured. The limitations are little effective depth, excessive natural drainage, seasonal droughts and potential soil blowing before practices of incorrect handling of the soil. The moisture regime is ustic and of moderate temperature (Cano *et al.* 1980). The current land use is primarily agricultural-livestock mixed type and main harvest crops are wheat, sunflower, corn, soybean and sorghum (Lorda, H *et al.* 2008).

2.2. Data Field

The study area was determined based on field data collection from the National Agricultural Information Network (RIAN) from National Institute of Agricultural Technology (INTA), Agricultural Experiment Station (EEA) Anguil. They have three transects of 10 km each, shape format with description of land use for batch and the corresponding surface in four different periods in November and April 2011 and November and April 2012, for the Capital department, province of La Pampa. For each transect, depending on the month, the following observations are made based also on the available satellite images:

**APRIL**

- **Winter Crops**: forage that are exploited only a few months, so we considered "annual". They are implanted from February to June (depending on the species) and graze until late spring (oats, barley, rye, wheat forage).
- **Sunflower**: On this date already harvested. It identifies the recent crop stubble.
- **Soybeans**: In April, this crop is in "maturity" and as stubble.
- **Perennial Pasture**: Pasture Alfalfa pure or mixed with grasses. They are implemented and are producing for 3, 4 or 5 years.
- **Plot**: Lots without crop residues (except sunflower), undifferentiated lots, lots plows, highly degraded pastures close to plowing (not found in forage production deficit).
- **Sorghum**: On this date grain and forage sorghum can be well differentiated.
- **Corn**: Normally developed, some in full maturity.
- **Peanuts**: This crop can be either freshly harvested or mature.
- **Natural Field**: Natural grass which are not changed through the years. In April, November and December looks the same.
- **Weeping Grass**: Pasture permanent once implanted can last 30 years. It looks the same in April and November – December.

**NOVEMBER-DECEMBER**

- **Winter crops**: Some are already grazed or harvested, others in production (wheat, oats, barley and rye).
- **Sunflower**: Found in implementation.
- **Soybeans**: Also on this date in implementation but not yet finished planting.
- **Perennial Pasture**: They look like in April, as they are always leaf.
- **Plot**: No difference from April
- **Corn / sorghum**: Unable differentiate grain sorghum from forage sorghum or maize.
- **Peanuts**: This culture may be in implementation.
- **Natural Field**: There is no difference with the month of April.
- **Weeping Grass**: There is no difference with the month of April.

2.3. Landsat data

Landsat 5 was launched in 1984 and continued to acquire imagery. Landsat’s strengths are generally seen to be its regular acquisition schedule (revisits the same area every 16 days), long-term data archive (image with comparable specifications is available from 1982), and relatively rich spectral information. Limitations of Landsat data are that it is only a moderate-resolution image source (30m multispectral data, 120m thermal infrared), and the fixed acquisition schedule makes it sometimes difficult to acquire imagery for a particular place at a particular time. From October 2008, all Landsat TM archived imagery and new
acquisitions are free and can be downloading from the USGS Global Visualization Viewer (GLOVIS), the USGS http://earthexplorer.usgs.gov/ or Landsat.org. (http://landsat.usgs.gov). The image used in this work corresponds at the Path/Row 228/085 Recently, the NASA set in orbit the Landsat 8 that will serve to monitor and improve the results of classification of use of soil and that they are already free available. Landsat 8 providing moderate-resolution imagery, from 15 meters to 100 meters, of Earth’s land surface and polar regions, Landsat 8 will operate in the visible, near-infrared, short wave infrared, and thermal infrared spectrums. Landsat 8 will capture approximately 400 scenes a day (http://www.nasa.gov/).

2.4. ALOS PALSAR data

The Phased Array type L-band Synthetic Aperture Radar (PALSAR) is an active microwave sensor using L-band frequency to achieve cloud-free and day-and-night land observation PALSAR will have ScanSAR, which will enable us to acquire a 250 to 350km width of SAR images (depending on the number of scans) at the expense of spatial resolution. This swath is three to five times wider than conventional SAR images (http://www.eorc.jaxa.jp/ALOS/en/about/palsar.html)

To realize the analysis, there was requested the image nearest to the date of the information that turned out to be of September 18, 2010, in FBD (Beam Double Polarisation Dies) mode and with a processing level 1.0.

3. Materials and Methods

In this paper, the methodology used for the classification of Landsat 5 TM images, was based on the grouping of various crops as the spectral signature of these and the confusion matrix obtained in the post classification with the software ENVI 4.3.

In figure 1 can be observed how the spectral signature of the cultivation relieved to field in November 2011, they have the same tendency and are confused with each other, except for the correspondents to the classes Soil (yellow) and Water (blue).

Figure 1: Spectral firms of crop, November 2011. Department Capital La Pampa.

After a first approximation with ISODATA unsupervised classification, we used three supervised classifiers. There were elected the most popular classifiers and that, in the last years, are the object of numerous studies on their performance in this field (Shao and Lunetta, 2012; Kavzoglu and Colkesen, 2009; Pizarro et.at, 2012; Pratola et al, 2012).

Maximum Likelihood classifier is a parametric method, that is to say, it assumes that the statistics for each class in each band is normally distributed and calculates the probability that a certain pixel belongs to a certain class (Schowengerdt, 2007). Two nonparametric methods, Neural Net and Support Vector Machine (SVM), were also considered, that do not need of supposed statistics about the distribution of probability of the information.

3.1. Unsupervised classification

ISODATA was employed for a first classification of the study area to be able to identify the spectral cluster associated to each land cover type independently from any prior information.

3.2. Selection of Region of Interest (ROIs)

With the purpose of identify the spectral signatures of crops relieved in the fields (November 2011), a selection of the Region of Interest (ROI) was performed. For each crop pixels for training and their corresponding test were sampled.

The categories were compared with the ISODATA unsupervised classification and field data. From this classification, different signatures of the same crop were separated and obtained more than one class of sunflower (Sunflower 1, 2 y 3), corn (Corn 1, 2 y 3) and winter crops (Winter crops 1 y 2) which, together with the other classes gave a total of 18 types of crops. Also, the classes such as urban, forest, soil and water were extracted from the image.
3.3. Supervised Classification

The supervised classifications was carried out using the ROIs previously obtained by exploiting three types of classifiers: Maximum Likelihood, Neural Net and Support Vector Machine.

3.3.1. Maximum Likelihood

With the results of the unsupervised classification the supervised classification Maximum Likelihood using ENVI 4.8 software was used. In order to analyze the quality of the obtained classification, we calculated the corresponding confusion matrix. This resulted in a global accuracy of 47.04%, while the coefficient Kappa obtained was 0.42.

3.3.2. Neural Net

In the same way that in the former case, the confusion matrix was calculated of the Neural Net. The global accuracy was the 57.15% and the coefficient Kappa of 0.5. While this classifier was found to be more accurate than the previous one, both indicate that classes are not really separable.

3.3.3. Support Vector Machine

The number of pixels correctly classified by Support Vector Machine classifier was inferior to the previous classifiers (44.67% accuracy) as well as the coefficient Kappa (0.39).

4. Results and discussion

4.1. Mono-temporal

To improve the results we proceeded to the generation of clusters, joining the training classes that were not separable. In each test, the Neural Net classifier determined greater accuracy in the classification (Table 1).

<table>
<thead>
<tr>
<th>Classes proposed</th>
<th>Categories that form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>Water</td>
</tr>
<tr>
<td>Forest</td>
<td>Forest, Caldén, Natural Fiel</td>
</tr>
<tr>
<td>Soil</td>
<td>Soil</td>
</tr>
<tr>
<td>Sunflower</td>
<td>Sunflower1</td>
</tr>
<tr>
<td>Corn</td>
<td>Corn1 y Corn3</td>
</tr>
<tr>
<td>Wheat</td>
<td>Wheat</td>
</tr>
<tr>
<td>City</td>
<td>City</td>
</tr>
</tbody>
</table>

Table 1: Cluster of crop identified in the classification process.

To obtain a clearer classified image, a Majority filter was applied with a window of 3x3 with the purpose of replacing the scattered pixels of another class within a larger class. Of each 9 pixels considered, the predominant value or class name is assigned to the center pixel in the output map. If no predominant value is found, the value or class name which is encountered first is used as output.

From successive Neural Net tests, eight classes were obtained. The overall accuracy (77.59 %) and Kappa coefficient (0.72) (Figure 2).

![Confusion matrix mono temporal image](image)

Figure 2: Confusion matrix mono temporal image.

It is observed that the classification determined a good result for the cultivation of sunflower, corn and wheat. Although in the first instance this class was divided in 3 different classes, the confusion matrix showed that some of these classes were confused with the class plot.

The confusion and differentiation both for the sunflower and for the corn, it can be due to the fact that the plant can be in phenological stage diverse since in this epoch newly its implantation begins. The wheat, on the contrary, would be in full development.
4.2. NDVI

The Normalized Index of Vegetation (NDVI) is based on the fact that the vegetation has a very high reflectivity in band 4 (NIR) of Landsat and very low in the band 3 (RED). Therefore, the greater the difference between bands larger the percentage of vegetation cover and healthier is the vegetation (Chuvieco, 2006).

This index was calculated for Landsat images for the months of February, March, April, October and November 2011 and the spectral signatures for each crop were extracted in order to analyze the annual difference between crops (Graph 2). As in Figure 3, observed confusion among different crops. In this case, the signatures of water, the city, the sunflower 3 and winter crop 1 can be differentiated.

![Figure 3: NDVI for all Landsat images from 2011.](image)

4.3. Multitemporal

To observe the behavior of crops classified, was performed a multi-temporal analysis with Landsat images of the months of February, March, April, October and November 2011. The overall accuracy for classification Neural Net of the eight classes set forth above, determined a value of 85.41% while Kappa coefficient was 0.82 (Figure 2), resulting in an improvement of the classification of the groups of crops.

![Figure 2: Confusion matrix multitemporal image.](image)

In order to observe the frequency distribution of the values of reflectance of some crops, Sorghum G and Corn1 histograms of the bands 4 were extracted.

The Bands 4 of the Landsat 5 TM satellite corresponds to the near infrared and is useful for determining biomass since the reflectivity is very high because healthy vegetation absorbs little energy in this band (Chuvieco, 2006).

In figure 4 you can see the histogram for the cultivation of Sorghum G. This indicates that higher levels reflectance of this crops occur in the months of February, March and April which explains why in the analysis of the image of the month of November, the same could not be classified.

![Figure 4: Sorghum G. histograms. Multitemporal images 2011.](image)

In contrast, in figure 5, we can see how the corn1 crop has its highest levels of reflectance in the months of November and February.

![Figure 5: Corn1 histograms. Multitemporal images 2011.](image)

In the other hand, in Figures 6 and 7 the classifications for the layer stacking of all year 2011 and November 2011, respectively are shown. With this comparison we can study changes in land use throughout the year in the study area.
In this case, it should be noted, that the changes in the City class are due to the presence of clouds in one of the images.

In the Figure 8, an enlargement of a region of the study area that suffered a big change going from classes Forest to Wheat in the month of November is illustrated.

References


Bentancourt, G. Las Maquinas de Soporte Vectorial (SVMs). *Scientia et Technica*, 27, 2005. ISSN 0122-1701.


