



This document offers a brief description about the project described.

USE OF SATELLITE IMAGES TO CALCULATE STATISTICS ON LAND COVER AND LAND USE

ABSTRACT

DANE has conducted a pilot project with the aim to propose a method to calculate the indicator 68 "Ratio of land consumption rate and population growth rate", which is part of the Sustainable Development Goals (United Nations, 2015). The methodology is based on remote sensing processing and GIS analysis; it was applied to the Barranquilla metropolitan area, through satellite images and population data for the years 2005, 2010 and 2015.

CONTENT

| | |
|--|----|
| 1. INTRODUCTION | 3 |
| 2. STUDY AREA | 4 |
| 3. DATA SOURCES | 5 |
| 3.1 Satellite imagery data | 4 |
| 3.2 Population data..... | 5 |
| 4. METHODOLOGY | 6 |
| 4.1 Imagery pre-processing | 6 |
| 4.2 Change detection method | 7 |
| 4.2.1 Image classification..... | 7 |
| 4.2.2 Definition and implementation of edition rules | 13 |
| 4.2.3 Accuracy assesment..... | 15 |
| 4.2.4 Change detection matrix..... | 16 |
| 4.3 Análisis demográfico | 16 |
| 4.4 Indicator estimation..... | 17 |
| 4.4.1 Population growth rate..... | 17 |
| 4.4.2 Land consumption rate | 17 |
| 4.4.3 Indicator estimation..... | 17 |
| 5. RESULTS..... | 17 |
| 5.1 Change detection | 17 |
| 5.2 Accuracy assesment..... | 21 |
| 5.3 Change Detection..... | 22 |
| 5.4 Demographic analysis: | 23 |
| 5.5 Rates computation..... | 25 |
| 6. RECOMMENDATIONS AND FUTURE WORK | 27 |
| 6.1 Data..... | 27 |
| 6.2 Methodology..... | 27 |
| 6.3 Training and skills..... | 28 |
| 6.4 Cost-Benefit Analysis | 28 |
| 6.5 Moving forward | 28 |

1. INTRODUCTION

As proposed in the Terms of Reference of the Task Team on Satellite Imagery and Geo-Spatial Data, the National Statistics Offices (NSO) are interested the use of satellite images to improve and eventually produce official statistics on a wide range of topics, such as the Sustainable Development Goals (SDG), and to promote capacity building and sharing of experiences.

To meet this requirement, DANE (The Colombia National Statistics Office) must incorporate new sources of information different from traditional, such as those offered by Big Data, specifically those from satellite imagery.

In this context, DANE has conducted a pilot project with the aim to propose a method to calculate the indicator 68 "Ratio of land consumption rate and population growth rate", which is part of the Sustainable Development Goals (United Nations, 2015), and seeks to obtain an estimation of land use efficiency, through monitoring the relationship between land consumption and population growth (SDSN, 2016):

"(...) this indicator benchmarks and monitors the relationship between land consumption and population growth. It informs and enables decision-makers to track and manage urban growth at multiple scales and enhances their ability to promote land use efficiency. In sum, it ensures that the SDGs address the wider dimensions of space and land adequately and provides the frame for the implementation of several other goals, notably health, food security, energy and climate change".

According to this definition, the objectives of the project are:

- To assess the feasibility of applying remote sensing methods for the estimation of land consumption rate, and Geographic Information Systems (GIS) based spatial analysis for the calculation of the relationship between land consumption and population growth in metropolitan areas.
- To evaluate the methodology by its application to the Barranquilla metropolitan area, through satellite images and population data for the years 2005, 2010 and 2015.

2. STUDY AREA

The initial disaggregation level proposed by UN for the indicator includes metropolitan areas (MA). In particular, Colombia has five MA that have been identified technically and legally, being Barranquilla MA among them. Keeping in mind this situation and the consideration on the availability of satellite images that satisfied the design of the project, this city and its MA was chosen.

The Barranquilla MA is located in the North part of Colombia. It consists of Barranquilla, officially the municipality core and capital of the Atlantic Department, and the municipalities Galapa, Malambo, Puerto Colombia, and Soledad (Fig. 1). Its area is 512 km² and the population is 1726271 inhabitants, according to the 2005 census. It occupies part of the Hydrographic Basin of the Mallorquín Swamp and a portion of the Wetlands Complex of the Magdalena River (CORMAGDALENA; CRA; DAMB, 2006), (Atlántico, CORMAGDALENA, & Internacional, 2007)

The Area has an undulating relief (west) to flat (west bank of the Magdalena River) for the most part. Hilly relief in some areas is located in the downtown district of Barranquilla with gradual inclination terrain to the Magdalena River, east of the city. The climate is generally tropical. The total mean annual rainfall stands at 815 mm and the average annual temperature is kept constant ranging from 27,5 °C on the strip coastal (Caribbean Sea) and 27,7 °C toward the south.

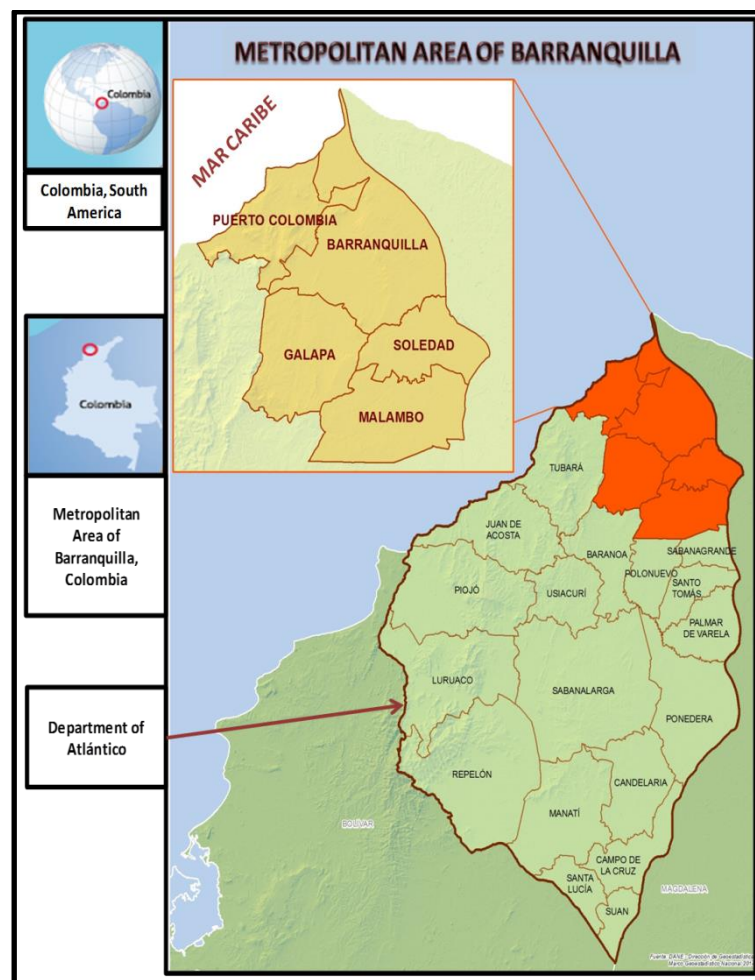


Figure 1. Barranquilla MA: Location map and municipal composition

3. DATA SOURCES

1.1. Satellite imagery data

To calculate the indicator, the time interval of five years was chosen. This criterion was considered appropriate to see the evolution in the population and to have adequate imagery coverage. The years included in the analysis are 2005, 2010 and 2015.

The criteria to choose the images were:

- Availability.
- Cloud coverage under 10% of the study area
- Same month for each year to minimize the changes generated by rainy or drought seasons, phases of crops and others.
- Full coverage of the metropolitan area to avoid image mosaics.
- Similar spectral bands.

Free Landsat images were used. They were downloaded from the United States Geological Survey (USGS) Earth Explorer website (<http://earthexplorer.usgs.gov/>). The study area is covered by the Landsat path-row 9-52. In the table Table 1 are shown for each image: mission, date and filename.

| Path-Row | Mission | Capture date | File |
|----------|-----------|-----------------|-----------------------|
| 9-52 | Landsat 7 | January 7 2005 | LE70090522005007ASN00 |
| 9-52 | Landsat 7 | February 8 2005 | LE70090522005039EDC00 |
| 9-52 | Landsat 5 | January 29 2010 | LT50090522010029CHM01 |
| 9-52 | Landsat 8 | January 11 2015 | LC80090522015027LGN00 |

Table 1. Landsat images included in the project.

For 2005, two scenes were collected. This was necessary to perform an enhancement on the image. The process is described in the pre-processing section.

3.1 Population data

Spite of census population data are available for 2015, for the years 2010 and 2015, projected data have been calculated by DANE with the demographic components statistical method, as well as cohort relation (DANE, 2016). In order to make the population data comparable for the three years, it was necessary to use also the projected data for 2005.

4. METHODOLOGY

In Figure 2 are depicted the main phases and procedures followed in the project implementation.

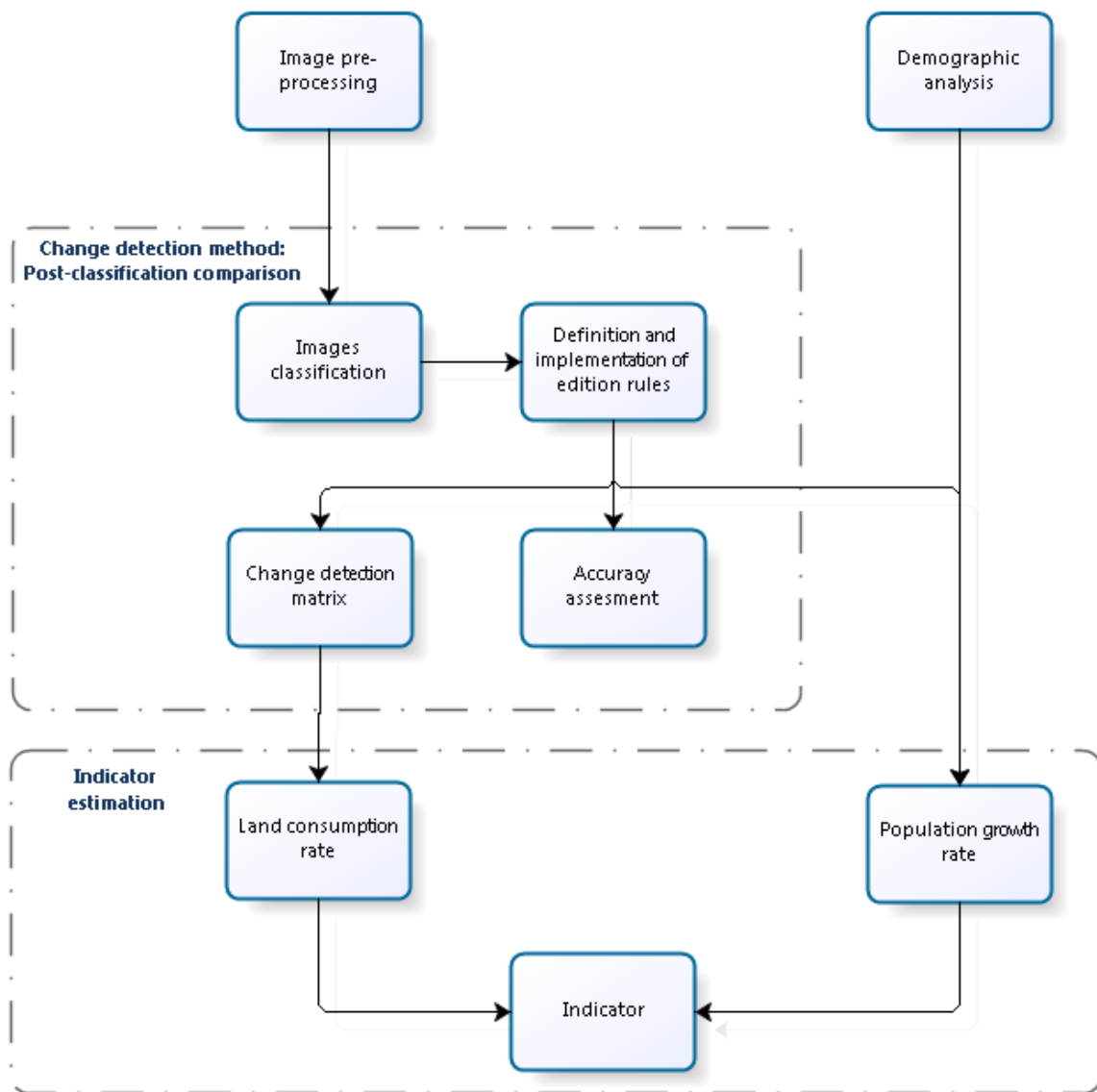


Figure 2. Flow chart for the computation of the SDG Indicator 68

4.1. Imagery pre-processing

This step includes the procedures carried out in order to prepare the images for the classification process. : filling gaps (for the year 2005 image), radiometric adjust (for the year 2015 image), cartographic projection, subset image creation and stack bands.

4.1.1. Filling the gaps

Landsat images available for 2005 were captured by the Landsat 7 ETM+ sensor, and they show data gaps due to lack of Scanning Line Corrector (SLC). According to Landsat Missions Web site *“there are methods available that allow users to fill the gaps of Landsat 7 data to create an aesthetic nice-*

looking image and methods that better maintain the integrity of the data, to create an image better suited for scientific interpretation and analysis”.

The USGS Gap-Fill Algorithm was used to improve the 2005 images. In general terms, it works as follows: The algorithm receives as input two scenes corresponding to the same band of the images to be corrected, primary and secondary (or filling), and returns a scene whose valid data will be the same as the primal scene attached to the gaps that may have been filled . The steps are:

- Around the pixel X, Y with no data is in the primary image, looks for a window that is scaled by the neighbors that surround it, until it reaches the minimum amount required for common pixels (valid pixels in both images), 144 suggested by USGS.
- Estimates the gain and bias using the mean and standard deviation of the common pixels found in the primary image and the secondary.
- Calculates the prediction from the function: Value Pixel Imagen Secondary filled * gain + bias.
- Assigns the value to the original pixel that had no data.

Considering that eliminating gaps in the pixels that are on the edges could lead to a prediction bias, it was decided to expand the area of interest in 1 000 m around the Barranquilla MA. This algorithm was programmed R software (R Foundation, 2015).

Since ERDAS software was available, it was used to convert from image to plane txt. See Annex 1.

4.1.2. Cartographic projection

With this process consistent plane coordinates were assigned to the images in order to allow the calculation of areas and perimeters during the interpretation procedure.

Landsat images are originally referred to WGS84 reference system and UTM plane coordinates. To make them compatible with the national reference system of Colombia, SIRGAS reference system and Transverse Mercator plane coordinates were assigned with the geoprocess routine.

For the year 2005, the filling-gap process led to a plane file used to generate the .img format image. The same geo-referencing procedure was applied.

4.1.3. Radiometric resolution scaling

The Landsat 8 image for 2015 has a radiometric resolution of 10 bits, and the 2005 and 2010 images have an 8 bits value. Since it is desirable to have all the images in the same rank of digital levels, a scaling radiometric resolution procedure was applied to the 2015 image to ensure the same radiometric resolution, i. e. 8 bits.

4.1.4. Subset image creation

Subsets of the images were generated covering a buffer of 500 meters around the metropolitan area of Barranquilla. This distance is a compromise between the coverage of the surrounding areas and the spatial limitation of the 2010, Landsat 5 image, which is smaller than those from the other missions.

4.1.5. Layer stacking

The transition from Landsat 5 to Landsat 7 and then to Landsat 8 included the addition of new bands and, some previously defined were sub-divided. Even so, the most of them are comparable. Table 2 shows the band classification employed.

| Order | Name band | Landsat 5 and 7 | Landsat 8 |
|-------|------------------------------|-----------------|-----------|
| 1 | Blue | Band 1 | Band 2 |
| 2 | Green | Band 2 | Band 3 |
| 3 | Red | Band 3 | Band 4 |
| 4 | Near Infrared (NIR) | Band 4 | Band 5 |
| 5 | Short Wave Infrared (SWIR) 1 | Band 5 | Band 6 |
| 6 | Short Wave Infrared (SWIR) 2 | Band 7 | Band 7 |

Table 2. Layer stacking done for the Landsat images.

4.2. Change detection method

Since the goal was to identify changes in land cover for three time periods, change detection techniques were required. Techniques commonly used for change detection are image differencing, image overlay, image index, image regression, post-classification comparison, among others.

Jensen, 1993, cited by (Battha, 2010), the post-classification comparison is currently the most popular method of urban change detection. In this method, each date of rectified imagery is independently classified to fit a common land type schema (equal number and type of land-cover classes). The resulting land cover maps are then overlaid and compared on a pixel-by-pixel basis. The result is a map of land-cover change. This per-pixel comparison can also be summarized in a ‘from-to’ change matrix, also called transition matrix (Jensen, 2005). The ‘from-to’ change matrix shows every possible land cover change under the original classification schema.

4.2.1. Image classification

Image classification is the process used to produce thematic maps from imagery (Battha, 2010); (Schowengerdt, 2007), (Jensen, 2005). There are basically two methods for image classification: First, Pixel-based, in which each one is classified depending on its spectral characteristics in a single class; and Object-based Image Analysis -GEOBIA¹- that analyze both: the spectral and spatial/contextual properties of pixels, then aggregates image pixels into spectrally homogenous image objects using an image segmentation algorithm and finally classifies the individual objects (Riggan & Weih, 2010), (Xia & Liu, 2010), (Costa et al, 2013).

¹ **OBIA** stands for Object-Based Image Analysis. Some authors argue that geographic space is intrinsic to this analysis, and as such, should be included in the name of the concept and, consequently, in the abbreviation: “Geographic Object-Based Image Analysis” - **GEOBIA**-. **Fuente especificada no válida**. In this document, the last definition is preferred

According to the literature review, GEOBIA has a number of strengths and weaknesses when compared to the conventional pixel classification see Table 3.

| Strengths | Weaknesses |
|---|--|
| <p>Information filter: GEOBIA is able to filter out meaningless information and assimilate other pieces of information into a single object. (Gronemeyer, 2015).</p> <p>Using image-objects as basic units reduces computational classifier load by orders of magnitude, and at the same time enables the user to take advantage of more complex techniques (e.g. non-parametric). (Hay & Castilla, 2006).</p> <p>Image-objects exhibit useful features (e.g. shape, texture, context relations with other objects) that single pixels lack. (Hay & Castilla, 2006)</p> <p>Provide classification results in a form that is immediately useable in a geographic information system (GIS; Geneletti & Gorte 2003, citado en (Frohn, Autrey, Lanes, & Reif, 2008).</p> <p>It reduces within-class spectral variation, and it removes the common salt-and-pepper effect that results from a pixel-based classification. (Xia & Liu , 2010)</p> <p>Multiple scales: The spatial relationship information contained in image objects allow for more than one level of analysis. (Gronemeyer, 2015)</p> <p>Accuracy: provide classification results with higher accuracy (Stuckens et al. 2000, Geneletti & Gorte 2003, citado en (Frohn, Autrey, Lanes, & Reif, 2008).</p> | <p>Under the guise of ‘flexibility’ current commercial object-based software provides overly complicated options. (Hay & Castilla, 2006)</p> <p>There are numerous challenges involved in processing very large datasets. Even if GEOBIA is more efficient than pixel-based approaches, segmenting a multispectral image of several tens of mega-pixels is a formidable task (efficient tiling/multiprocessing solutions are necessary). (Hay & Castilla, 2006).</p> <p>Segmentation is an ill-posed problem, in the sense it has no unique solution, e.g., (i) changing the bit depth of your heterogeneity measure can lead to different segmentations. (ii) Remember, even human photo-interpreters will not delineate exactly the same things. (Hay & Castilla, 2006).</p> <p>There is a lack of consensus and research on the conceptual foundations of this new paradigm, i.e., on the relationship between image-objects (segments) and landscape objects (patches). For example, (i) what is the basis to believe that segmentation-derived objects are fine representations of landscape structural-functional units? (ii) How do you know when your segmentation is good? (iii) Is there a formally stated and accepted conceptual foundation? (Hay & Castilla, 2006).</p> |

Table 3. GEOBIA SW Analysis. Adapted from Hay & Castilla (2006).

Since one of the processes in the project is to identify urban settlements, but their spectral response is very similar to that from barren or minimal vegetation, the GEOBIA method was used in order to include additional features to the spectral ones and to improve the thematic accuracy. To perform the process, the open-source INTERIMAGE- was used (no license payment).

Figure 3 describes the components of the interpretation process on INTERIMAGE. The system implements a specific interpretation control strategy, guided by a structured knowledge model

through a semantic net. The interpretation control is executed by the system core, which uses as input a set of geo-referenced images, SIG layers, digital elevation data or other geo-registered data. Through the interpretation of the scene, input data are processed with the help of external programs, called top-down and bottom-up operators.

Top-down operators are responsible for the partition of the scene into regions, considered as object hypotheses (segmentation). This is a preliminary classification which identifies segments with the potential to belong to each class. The bottom-up operators refine the classifications produced in the top-down step (confirming or rejecting) and solving possible spatial conflicts between them. At the end of the interpretation process, the hypotheses become validated object instances (classification).

The image to be classified, and the semantic network (hierarchical entities that will be identified in the image) are entered to INTERIMAGE as input. Then, the methods and operators that allow the generation of objects and their classification are defined.

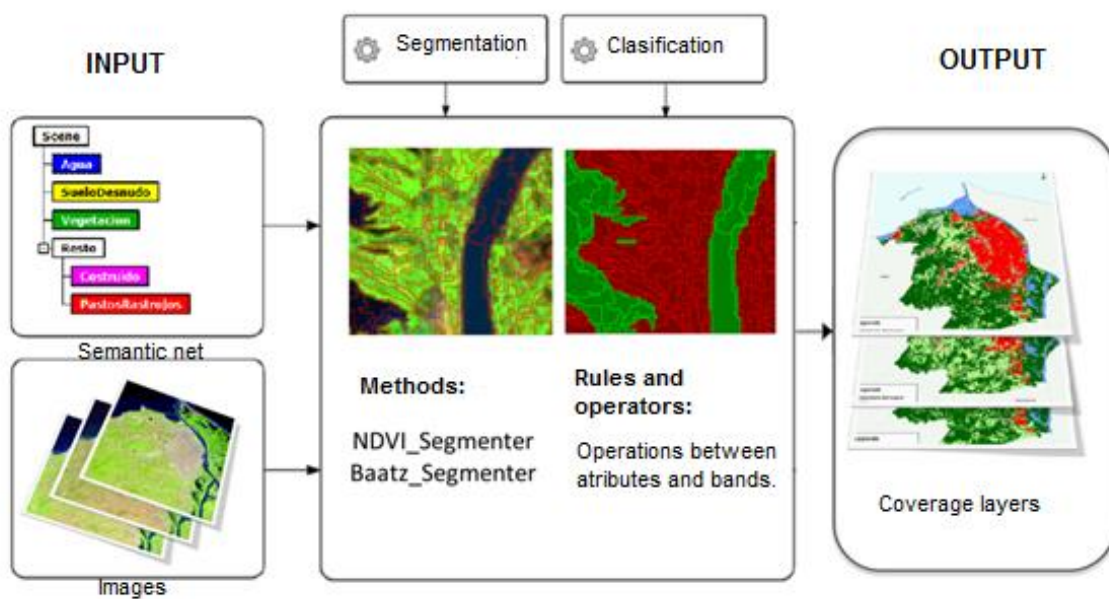


Figure 3. INTERIMAGE flow chart for object-based classification.

4.2.2. Semantic Net

Semantic net is a hierarchical structure of entities that are expected to be found in the image. Each node in the semantic net has properties, such as top-down and bottom-up operators as well as generic parameters and other specific operators.

The project's semantic net was created based on the previous knowledge of the study area and taking into account the Landsat imagery land cover classification from the project GeoCover². See Table 4:

| Coverage | Description |
|-------------------------|--|
| Water | All water bodies, artificial or natural. |
| Urban or built-up areas | Area developed with buildings and other civil works within a |

² <http://www.mdafederal.com/geocover>

| | |
|-------------------------------------|---|
| | defined urban perimeter. |
| Grassland | Upland herbaceous grasses |
| Barren or minimal vegetation | Land with minimal ability to support vegetation, including rock, sand, beaches and mines. |
| Forest evergreen and shrub or scrub | Species that do not seasonally lose leaves. Includes both broadleaf and needle leaf species, as well as evergreen tree species in the wetland environment. It also includes woody vegetation wetlands |

Table 4: Semantic Net entities included in the project

4.2.3. Image segmentation

A necessary prerequisite for object oriented image processing is the successful image segmentation (Baatz & Schape, 2000). It generates a set of non-overlapping segments/polygons that is expected to have relatively homogeneous and semantically significant groups of pixels (Blaschke, 2010, cited by (Costa et al, 2013)).

There are many types of segmentation algorithms which establish different criteria to create objects. The most common are growing technique in which smaller image objects are sequentially merged into bigger ones. Image segmentation starts with each pixel forming one image object or region.

At each step, a pair of image objects is merged into a larger object. The merging decision is based on local homogeneity criteria, describing the similarity of adjacent image objects. The procedure stops when there are no more possible merges because the increase on heterogeneity exceeds a defined threshold. (Baatz & Schape, 2000).

The image segmentation algorithms were:

NDVI Segmenter: This algorithm is based in the Normalized Difference Vegetation Index (NDVI). It is the normalized difference of green leaf scattering in near-infrared, and chlorophyll absorption in red. It is defined in Equation 1:

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

Where NIR and RED stand for the spectral reflectance measurements acquired in the visible red and near-infrared regions, respectively. NDVI segmenter allows identifying vegetation coverages. It requires as an input a threshold for the NDVI. The common range for green vegetation is 0,2 to 0,8.

Baatz segmenter: is based on an iterative process of local optimization that minimizes the mean heterogeneity inside each segment. The heterogeneity criterion is based on color and shape (Baatz & Schape, 2000), Equation 2:

$$f = w * h_{color} + (1 - w) * h_{shape} \quad (2)$$

Where, f expresses the increase of heterogeneity resulting from the union of two segments, before a union operation. The fusion factor is calculated for each of the neighbors of the selected segment. The neighbor which has the minimum fusion factor is chosen for merge.

However, the union only occurs if the fusion factor is under certain threshold, defined as the square of the scale parameter. This procedure continues merging segments until no more unions are possible. (Happ, Ferreira, Bentes, Costa, & Feitosa, 2010). The fusion factor contains a component for the spectral heterogeneity h_{color} , defined in equation 3 and a component for the spatial heterogeneity h_{shape} , defined in equation 4. The user-defined weight parameter w is based on the relevance of shape or color.

$$h_{color} = \sum_c w_c (n_{merge} * \sigma_c^{merge} * (n_{obj1} * \sigma_c^{obj1} - n_{obj2} * \sigma_c^{obj2})) \quad (3)$$

In Equation 3, n_{merge} is the number of pixels within a merged object n_{obj1} is the number of pixels in object 1; n_{obj2} is the number of pixels in object 2; σ_c is the standard deviation within object of band c ; subscripts merge refer to merged objects and $obj1$ and $obj2$ refer to the objects prior to a merge.

$$h_{shape} = w_{cmpct} * h_{cmpct} + (1 - w_{cmpct}) * h_{smooth} \quad (4)$$

Compactness degree is defined as a ratio between the segment perimeter and the squared root of its area in pixels units times the number of object pixels. Smoothness is defined as the ratio between segment perimeter and your minimum bounding box (Costa et al, 2013).

Initially, each segment (object) is defined as a single image pixel. The segment grow as they are merged with their neighbors always trying to minimize the incorporated heterogeneity and only if the minimum fusion factor f is less than an value e defined as a squared for a scale parameter. The process continues until any segment cannot grow more.

Table 5 shows spectral bands and algorithms that were used in the project. Bands were selected according to the combinations commonly applied to coverage identification.

| Coverage | Algorithm | bandas |
|-------------------------------------|-----------------|---|
| Water | Baatz segmenter | Near Infrared (NIR), Short Wave Infrared (SWIR) 1, Red |
| Barren or minimal vegetation | Baatz segmenter | Blue, Red, Short Wave Infrared (SWIR) 2 |
| Forest evergreen and shrub or scrub | NDVI | Red, Near Infrared (NIR) |
| Urban or built-up areas | Baatz segmenter | Blue, Near Infrared (NIR), Short Wave Infrared (SWIR) 1 |
| Grassland and shrub or scrub | N/A | N/A |

Table 5. Segmentation algorithms used in the project

4.2.4. Decision rules

After defining the segmentation algorithms, decision rules were introduced for each node of the semantic network. These rules allow filtering, from the set of objects, those that belong to the class of interest.

In a decision rule, it is possible calculate a variety of attributes based on spectral values, shape, texture and topological characteristics of image segments. These attributes can be used to select objects within a set, with a user-defined threshold.

Objects that are not within a threshold are evaluated in a different class according to the order defined in the semantic network. If an object meet the criteria for more than one class, it is assigned to that with the higher reliability.

For those objects that did not meet the criteria defined for any of the classes, a new one was created: Grassland and shrub or scrub.

| Coverage | Reliability | Expression | Threshold | | |
|-------------------------------------|-------------|-------------------------------|-----------|----------|---------|
| | | | 2005 | 2010 | 2015 |
| Water | 0.4 | Band Mean Add (layer5,layer6) | < 50.16 | < 42.71 | < 12.94 |
| Barren or minimal vegetation | 0.3 | Mean (layer5) | ≥ 134.3 | ≥ 126.85 | ≥ 58.89 |
| | | Band Mean Div (Layer3,Layer1) | ≥ 1.165 | 0.75 | |
| Forest evergreen and shrub or scrub | 0.2 | TA_NDVI_Segmenter | 0.03 | 0.2 | 0.5 |
| Urban or built-up areas | 0.1 | Mean (layer1) | ≥ 90.013 | ≥ 87.67 | ≥ 34.30 |
| Grassland and shrub or scrub | N/A | N/A | N/A | N/A | N/A |

Table 6. Decision rules, operators and indicators

4.2.5. Definition and implementation of edition rules

Two editing rules were used in order to ensure that each built area in the classification had a corresponding population register.

Rural Scattered editing rule: As a result of the image classification the class “Urban or built-up areas” was obtained. It includes urban settlements, roads, recreational centers, industrial facilities located in rural areas, among others; Figure 4.

An additional edition rule was created to depurate in the new class those elements that are not located in an urban agglomeration. To do it, geographic information of population centres and urban areas that DANE consolidates for statistical operations was used and let to identify the “scattered rural” category.

Rural scattered areas are understood as the scattered (dispersed) distribution of housing and agricultural exploitations that do not have an addressed road network and normally are deprived of public services and/or other urban facilities.

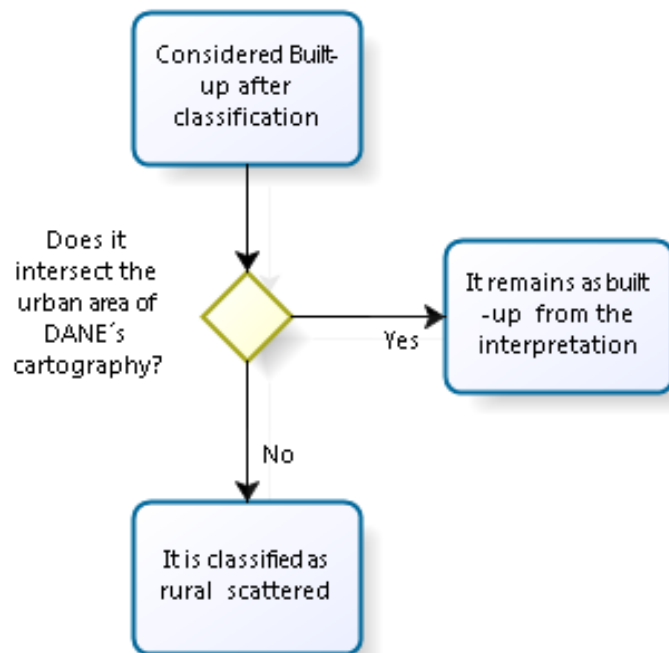


Figure 4. Rural scattered editing rule

Exclude Population or add polygon editing rule: Due to the spatial resolution of the satellite images, small population centres surrounded by tall vegetation are not easily detected. This led to use DANE maps as auxiliary reference. If a population center is not detected at any year, the population associated to it is excluded. This is done to keep the agreement between the consumption rate of soil and detected population growth rate.

When a population center is identified on DANE maps for the whole period, but not in the classification process, at any year, the cartographic information is used to complement the data about that center.

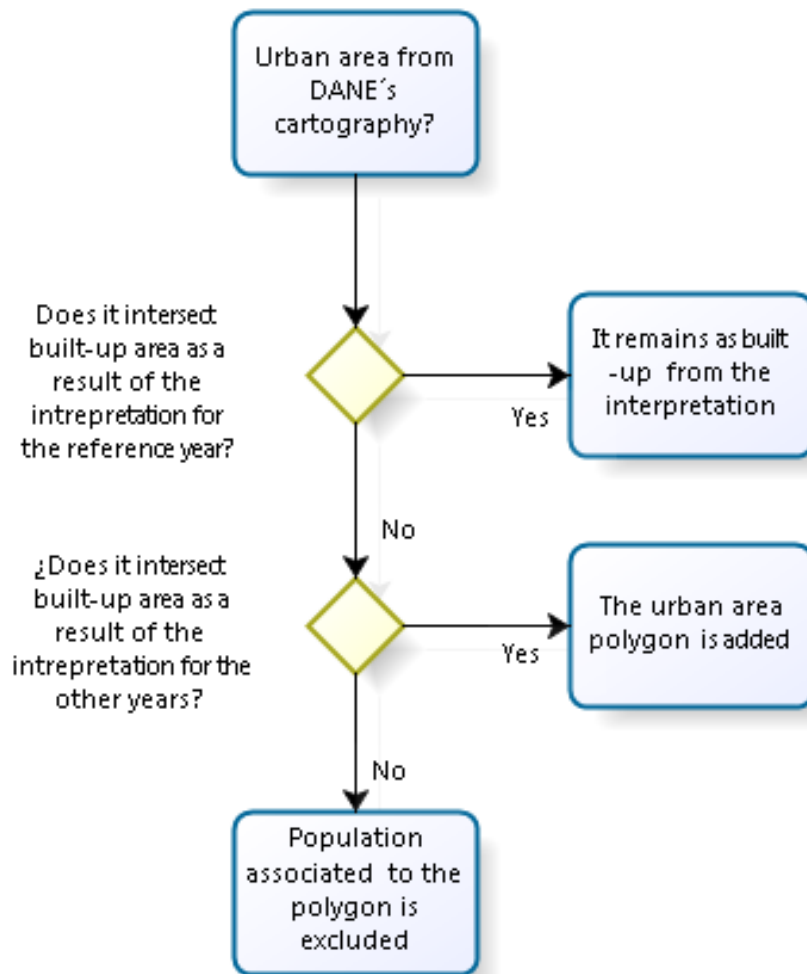


Figure 5. Exclude population or add polygon editing rule

4.2.6. Accuracy assessment

In the context of remote sensing classification, accuracy assessment is a measure of “*the agreement between a standard assumed to be correct and a classified image of unknown quality*” (Campbell, 2016).

A confusion matrix is a very effective tool to measure accuracy assessment because it compares two sources of information: pixels or polygons from a classification map developed from remotely sensed data and, ground reference test information (Jensen, 2005). The confusion matrix is the most common way to represent an assessment of the thematic accuracy classification map.

The confusion matrix was developed as follows:

- Determination of the sample size
- Distribution of the random sample of points
- Use of image and orthophotos as reference classes and assignation of a class for each point
- Check with automatic visual classification
- Accuracy computation

To determine the sample size, it was used the confidence interval estimation of similarity aggregated indicators as the proportion of successes in a binomial experiment (Rossiter, 2014):

$$n = \frac{Np(1-p)}{(N-1)\frac{d^2}{Z^2(1-\alpha/2)} + p(1-p)} \quad (5)$$

Where:

n is the sample size;

N is the number of pixels in the study area;

P is the population similarity a priori expected;

d is the ratio of relative sampling error and,

$Z_{(1-\alpha/2)}$ is the two-tailed score at the confidence level for p

The population size was defined by the number of pixels that cover the Barranquilla MA: N = 549866 pixels.

A priori accuracy was assumed as p = 87% and d = 10% (Ardila, Espejo & Herrera, 2005).

The Z score was calculated at a confidence level $\alpha = 0,95$, being $Z = 1,96$

The computed value for n is: n = 174 points.

After the random distribution of points, the join spatial operation allowed the assignation of each class, generated by INTERIMAGE, to each sample point and then it was evaluated the thematic accuracy of the obtained classification.

4.2.7. Change detection matrix

In order to quantify the changes in coverage according to the classification obtained in each of the years of study, it was performed a pixel by pixel comparison between two years of study, and was calculated the 'from-to' change matrix.

The 'from-to' change matrix shows every land-cover change under the original classification schema and shows the areas of change in each class. To do it, the resulting coverage maps were converted to raster format and processed in an algorithm in the R statistical software, which allowed for a change matrix for 2005-2010 and 2010-2015. (See annex 2).

4.3. Demographic analysis

An exploratory longitudinal data analysis, for the official population data of Barranquilla MA, was done taking into account de particular distribution for each municipality included. The Population Growth Rate (PGR) was estimated from the total aggregated population projections for the Barranquilla MA.

4.4. Indicator estimation

4.4.1. Population growth rate

To calculate the rate of population growth the following proposed formula was used:

$$PGR = \frac{Population(t_2) - Population(t_1)}{Population(t_1)} \quad (6)$$

4.4.2. Land Consumption Rate

The rate of land consumption was calculated from the resulting coverage maps according to the following proposed formula:

$$LCR = \frac{Urban\ Land\ Area(t_2) - Urban\ Land\ Area(t_1)}{Urban\ Land\ Area(t_1)} \quad (7)$$

4.4.3. Indicator estimation

The indicator is expressed as the ratio between land consumption rate and population growth rate.

$$Indicator(I) = \frac{Land\ Consumption\ Rate}{Population\ Growth\ Rate} \quad (8)$$

5. RESULTS

5.1. Change detection

5.1.1. Edition rules

After the image classification process, the edition rules were implemented on ArcGIS 10.1. Table 7 shows the areas belonging to the rural scattered category for each year:

| Rural scattered | 2005 | 2010 | 2015 |
|-----------------|--------|--------|--------|
| Area (ha) | 157,62 | 384,61 | 480,82 |

Table 7. Rural scattered category areas

As a result of the growth of mining and industrial infrastructure, as well as the building of new recreational centers in the northern part of the Barranquilla MA, the category rural scattered shows an increase during the period of observation.

With the rule “exclude population or add polygon” the population centers shown in the Table 8 were included from DANE maps, but they were not detected by the classification method:

| Municipal Code | Municipal Name | Population Center Code | Population Center Name | Population | | |
|----------------|-----------------|------------------------|--------------------------------------|------------|------|------|
| | | | | 2005 | 2010 | 2015 |
| 08296 | Galapa | 082962001001 | Paluato | 165 | 177 | 199 |
| 08573 | Puerto Colombia | 085732001004 | Urbanización Barranquilla Sport Club | 35 | 25 | 20 |
| 08758 | Soledad | 087582001001 | | 858 | 686 | 592 |

Table 8. Population centers not detected by the methodology

The population center Pitalito, at the Malambo municipality, was detected by the classification for 2015, but not for the other years: To include it in every year, its polygon (taken from DANE maps) was incorporated where necessary.

In this way, table 9 shows the population centers with maps and population records:

| Municipal Code | Name | Population Center Code | Population Center Name | Population | | |
|----------------|-----------------|------------------------|------------------------|------------|-------|-------|
| | | | | 2005 | 2010 | 2015 |
| 08001 | Barranquilla | 080012001001 | Pinar del Río | 3.876 | 3.624 | 3.657 |
| 08433 | Malambo | 084332001001 | Caracolí | 2.387 | 2.626 | 2.880 |
| 08433 | Malambo | 084332001004 | La Aguada | 753 | 828 | 908 |
| 08433 | Malambo | 084332001005 | Pitalito | 79 | 87 | 96 |
| 08573 | Puerto Colombia | 085732001002 | Salgar | 2.379 | 1.709 | 1.378 |
| 08573 | Puerto Colombia | 085732001003 | Sabanilla | 196 | 141 | 113 |
| 08573 | Puerto Colombia | 085732001008 | Villa Campestre | 2.101 | 1.509 | 1.216 |

Table 9. Population centers with available census maps and population records

This information was the basis for the subsequent demographic analysis and calculations of the rate of population growth for the Barranquilla MA. Finally, once implemented the SIG rules the following land cover results were obtained.

| CLASS | Area (ha) | | |
|-------------------------------------|-----------|-----------|-----------|
| | 2005 | 2010 | 2015 |
| Water | 4.014,63 | 3.725,31 | 3.934,25 |
| Urban or built-up areas | 9.882,49 | 10.636,64 | 11.231,32 |
| Grassland | 9.249,11 | 15.061,81 | 10.408,59 |
| Rural scattered | 157,62 | 384,61 | 379,54 |
| Barren or minimal vegetation | 531,63 | 444,93 | 764,56 |
| Forest evergreen and shrub or scrub | 25.656,02 | 19.238,20 | 22.773,50 |

Table 10. Land cover areas for the Barranquilla Metropolitan Area

In Figures 6, 7 and 8 is shown the land cover distribution for the years 2005, 2010 and 2015.

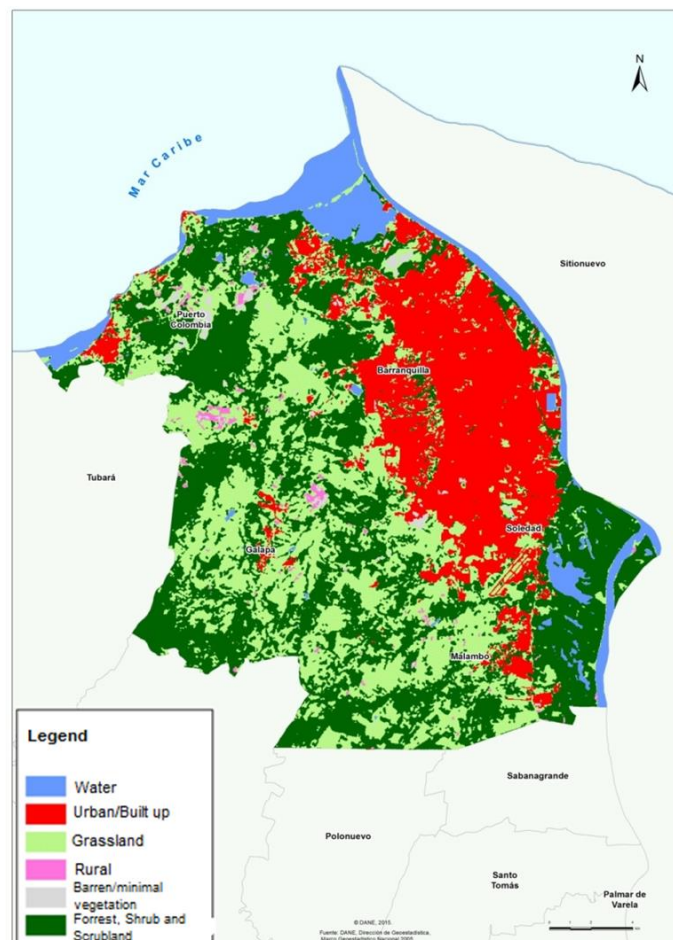


Figure 6. Land cover areas for the Barranquilla Metropolitan Area: year 2005

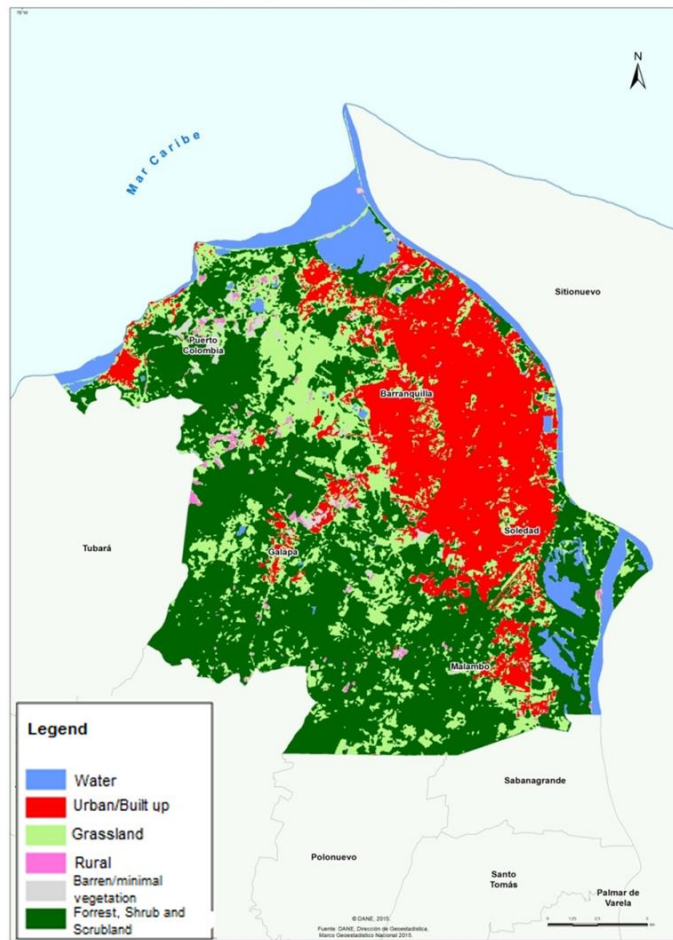


Figure 7. Land cover areas for the Barranquilla Metropolitan Area: year 2010

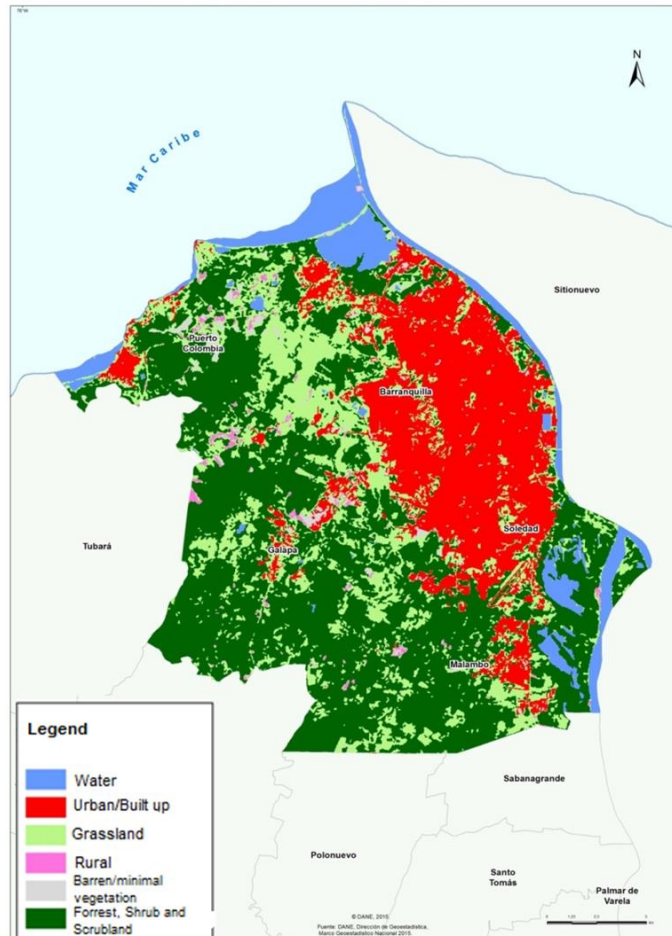


Figure 8. Land cover areas for the Barranquilla Metropolitan Area: year 2015.

From the above figures it could be observed that the grass and forest, shrub and scrubland are predominating for the three years. The changes in the grass cover (increase and further decrease) could be associated to El Niño event of 2009 – 2010 that implied for the northern coast of Colombia a rain deficit from 10% to 40% of the Normalized Difference Vegetation Index (NDVI).

The urban consolidated zones have been established close to the Magdalena River and the Caribbean Sea and it is observed a growth of the population centers in the central part of the Barranquilla MA.

5.2. Accuracy assessment

The confusion matrices contained in Table 11 were obtained for each year. Because of their representative behavior, the water, urban (built-up), grassland, forest and shrub or scrub classes have a good score:

| 2005 | Classification | | | | 2010 | Classification | | | | | 2015 | Classification | | | | | |
|----------------------|----------------|---------------|-----------|----------------------|----------------------|----------------|---------------|-----------|-------------|----------------------|----------------------|----------------|---------------|-----------|-----------------|-------------|----------------------|
| | Water | Built up area | Grassland | Forest and scrubland | | Water | Built up area | Grassland | Barren land | Forest and scrubland | | Water | Built up area | Grassland | Rural Scattered | Barren land | Forest and scrubland |
| Water | 85% | - | 15% | - | Water | 90% | - | 10% | - | - | Water | 90% | - | 10% | - | - | |
| Built up area | - | 95% | 5% | - | Built up area | - | 93% | 2% | - | 5% | Built up area | - | 83% | 15% | - | - | 2% |
| Grassland | - | - | 83% | 17% | Grassland | - | - | 86% | 2% | 13% | Grassland | - | 3% | 66% | - | 3% | 29% |
| Forest and scrubland | - | - | - | 100% | Barren land | - | 100% | - | - | - | Rural Scattered | - | - | - | 100% | - | - |
| Rural Scattered | - | - | 100% | - | Forest and scrubland | - | - | 3% | - | 97% | Barren land | - | - | 33% | - | 67% | - |
| | | | | | | | | | | | Forest and scrubland | - | - | 1% | - | - | 99% |

Table 11. Confusion matrices for the years 2005, 2010 and 2015

Visual and automatic comparison of matches and non-matches of classes assigned, allowed to determine the accuracy of the classification (Table 12). The measurements of the accuracy for each year are:

| Thematic | 2005 | 2010 | 2015 |
|-------------------------|-------|-------|-------|
| Number of matches | 161 | 159 | 151 |
| Number of No – matches | 13 | 15 | 23 |
| Precision percentage | 92,5% | 91,4% | 86,8% |
| Confidence interval 95% | 88,3% | 86,9% | 81,5% |
| | 97,6% | 96,7% | 92,9% |

Table 12. Classification accuracy

Although built areas have good classification accuracy, it is noted that, in some cases, these were classified as grasslands and scrubland. This could be explained by the similarity in the spectral response presented by both of the coverages.

Also, water bodies were susceptible to be classified as grasslands, this may be related to the presence of mangroves in the northern Magdalena River.

When forest areas had low density and/or poor health, they show NDVI values outside the preset range so they can become classified as grasslands. However, indicators of global accuracy are good for each of the three years.

5.3. Change Detection

After applying the post-classification and comparison methods, the changes detected were:

- Forest evergreen and shrub or scrub: from 52% to 46%.
- Urban or built-up areas: from 20% to 23%
- Grassland Barren or minimal vegetation: from 19% to 21%

Figure 9 presents the land coverage before built. It identifies that new urban areas have consumed 54% of land that was grassland before and 34% of forest evergreen and shrub or scrub.

Coverage land before being built up area in 2005 - 2015

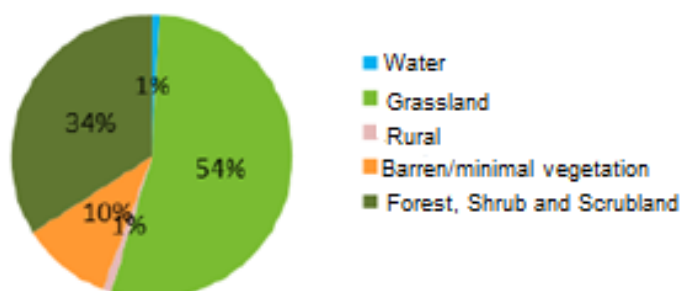


Figure 9. Land coverage before built-up from 2005 to 2015

5.4. Demographic analysis

According to the population projection, as well as the exclusion of population centers (to ensure the concordance of land consumption and population data), a demographic analysis was conducted.

The analysis shows that there were 3 528 707 inhabitants occupying this territory in 2005. Table 13 shows that the increase is lower in subsequent years.

| Year | Population | Class | | | Growth | |
|------|------------|----------------|--------------------|-----------------|---------|------------|
| | | Municipal town | Population centers | Rural scattered | Number | Percentage |
| 2005 | 3.528.707 | 1.747.745 | 12.829 | 8.617 | | |
| 2010 | 3.785.795 | 1.877.264 | 11.412 | 9.331 | 257.088 | 7,3% |
| 2015 | 4.039.936 | 2.004.617 | 11.059 | 9.395 | 254.141 | 6,7% |
| 2020 | 4.283.359 | 2.126.329 | 10.954 | 9.612 | 243.423 | 6,0% |

Table 13: Population changes in the Barranquilla metropolitan area for the years 2005, 2010 and 2015.

For the period 1985 - 2020, the population of Barranquilla shows a variation from 77,5% to 57,7% of the entire area, while Soledad changed from 15% to 32,7%. Moreover, Galapa, Malambo and Puerto Colombia municipalities show a linear growth and they include less than 10% of the group of municipalities.

This could mean that the most of the population growth in Barranquilla MA is mainly due to the contributions of Barranquilla city and Soledad (Table 14).

Growth rate between 2005 -2015
 Área Metropolitana de Barranquilla

| | Municipal Township | Population center | Urban area | Rural | Total |
|-----------------|--------------------|-------------------|--------------|-------------|--------------|
| Barranquilla | 6,3% | -5,7% | 6,2% | 230,4% * | 6,3% |
| Galapa | 35,2% | 20,6% | 35,2% | 19,7% | 33,6% |
| Malambo | 20,0% | 20,7% | 20,0% | 10,4% | 19,7% |
| Puerto Colombia | 8,7% | -42,1% | -0,7% | -21,4% | -2,6% |
| Soledad | 33,5% | -31,0% | 33,5% | | 33,3% |
| Total | 14,7% | -13,8% | 14,5% | 9,0% | 14,5% |

Table 14. Population growth rates

The rural population is about 5% of the total population over the study period. This shows an urbanization process in the Barranquilla MA.

In this study the urban area consists of municipalities and population centers that were detected in satellite images classification.

5.4.1. Population pyramids

In the figure 11 it can be seen that Soledad, Galapa and Malambo have a progressive structure with a population growth rate based on a high birth rate evidenced in the number of inhabitants between 0 and 20 years.

Furthermore, Barranquilla city and Puerto Colombia have a regressive structure since a great portion of their population is getting older and the birth rate keeps constant or trends to decline.

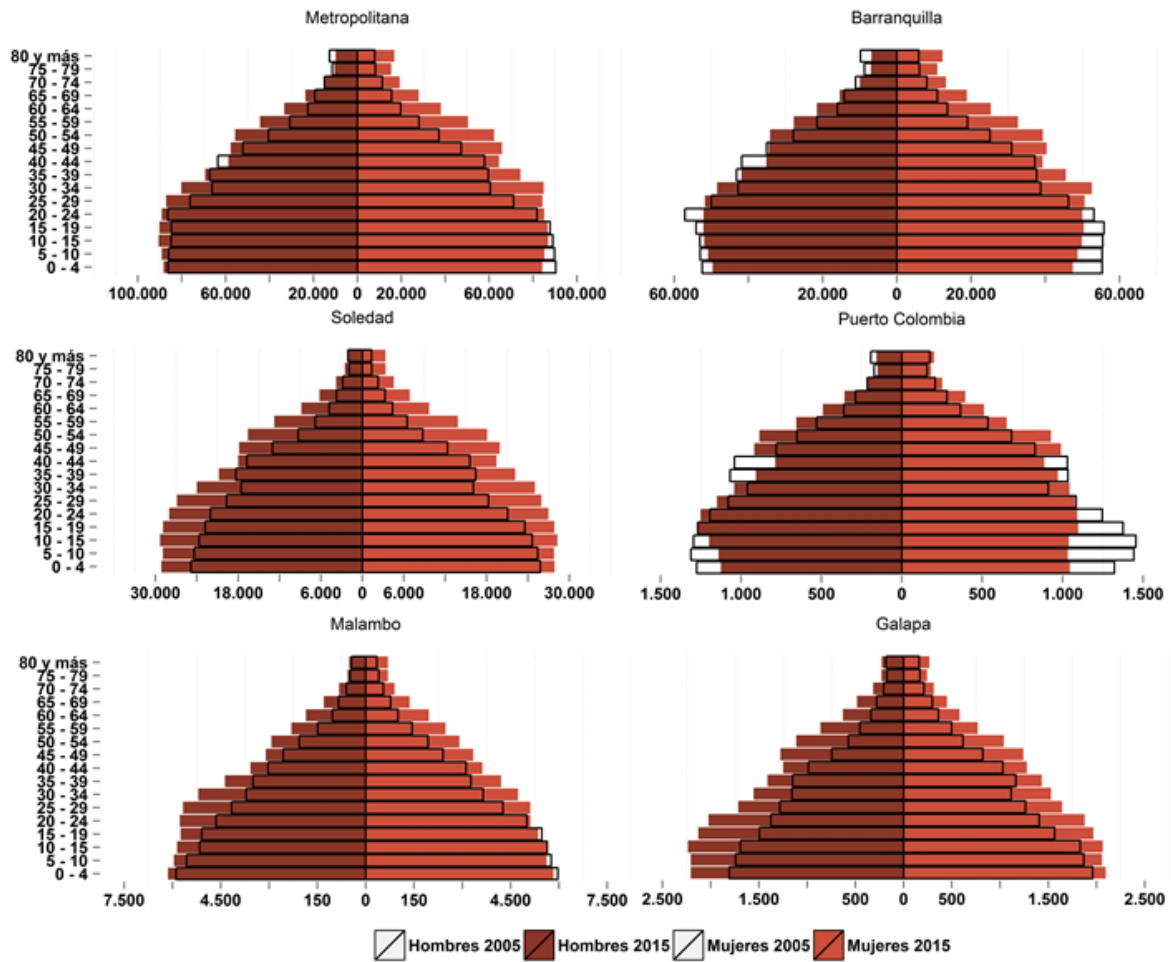


Figure 11. Population pyramids for the Barranquilla MA

Population is getting concentrated around the economically active age (29-53 years), which could mean that the Barranquilla MA acts as an attracting urban pole in which industry and services are mainly located.

5.5. Rates computation

From the classification procedure and the population data, Barranquilla MA has the following rates of growth.

| | 2005-2010 | 2010-2015 |
|-----------------------------|-----------|-----------|
| Urban land consumption rate | 8,1% | 4,7% |
| Population growth rate | 7,6% | 5,6% |

Table 15: urban land consumption and population growth rates

Therefore, the indicator that relates both rates is:

| | | |
|-----------|-----------|-----------|
| Indicator | 2005-2010 | 2010-2015 |
| | 1,06 | 0,84 |

Table 16: Ratio between urban land consumption and population growth rates

For the period 2005-2010 the indicator is greater than one, revealing a dispersed urban growth, i.e. the consumption of land is greater than the population growth. Therefore the population could be consuming land needed for agriculture and ecosystem services. For the period 2010-2015 trend is reversed (the indicator is lesser than one), revealing a possible process of urban grouping.

Municipal-level results show differentiated behaviors: Barranquilla and Galapa have land consumption greater than the population growth. Soledad and Malambo have indicators lesser than one, probably associated with a more efficient land use. See Figure 10.

Puerto Colombia has a particular behavior: It shows a loss of population but an increase in the land consumption. This can be associated to the construction of new facilities for recreational purposes.

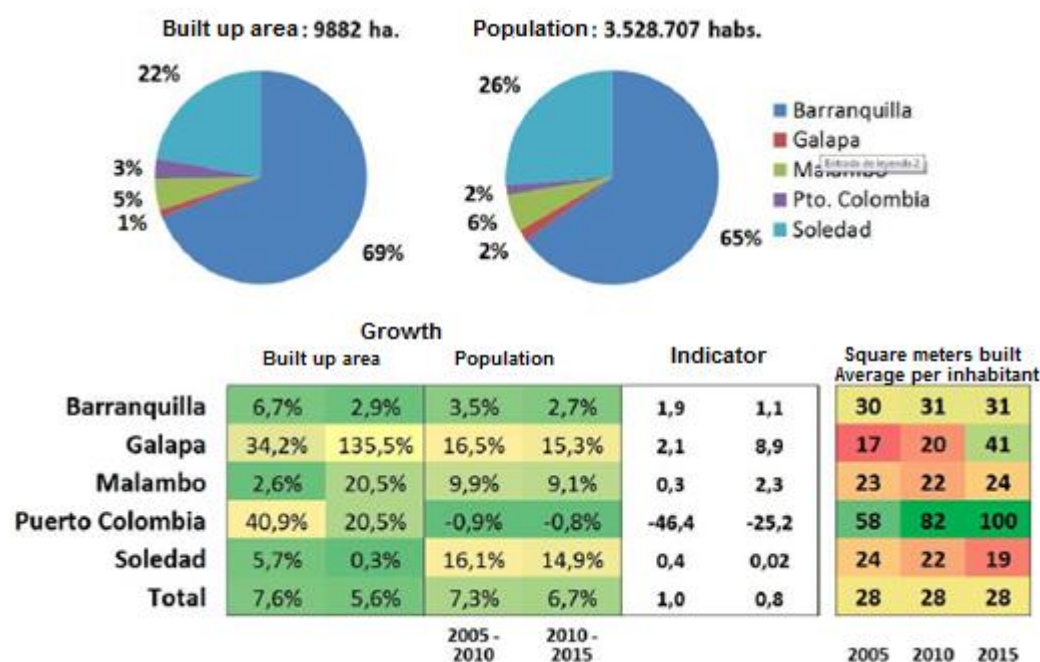


Figure 12. Evolution of the indicator and rates for the Barranquilla MA

6. RECOMMENDATIONS AND FUTURE WORK

6.1. Data

6.1.1. Geospatial data

Images with similar spatial and spectral resolutions, as well as availability for the same epoch during different years, are required for multi-temporal studies. These conditions reduce the quantity of images that can be used.

Landsat platform provides free access to satellite imagery data since 1971 but, in some cases, improvements must be done in order to overcome the cloud shadow and gaps issues. Though USGS documentation³ helps to solve these problems, the required tasks could be time consuming.

To exchange information and knowledge in a timely way, strategic partnerships promotion among national and international bodies, big data providers and academy are desirable.

6.1.2. Population data

Population projected data used in this exercise were obtained from the 1993 and 2005 censuses. This could mean accuracy limitations when used for the year 2015, particularly at small population centers and/or show different population growth rates.

Since they provide updated information, the administrative registers can help to solve this situation. Nevertheless, those registers are not initially intended for statistics. Their adequate transformation for the indicator computation is a challenge to the NSO's.

However, the advantages of administrative register, such as detailed measurement, local geographic levels and lower relative cost, make them a feasible complement in projects related to this type of indicators.

Additionally, registers must be georeferenced. This means the implementation of mechanisms for the capture based on location. Therefore, NSO's and administrative registers providers must conclude agreements on use, processing, custody, confidentiality and dissemination policies, among others.

6.2. Methodology

Validation is a mandatory procedure when Big Data are used for generation of official statistics. Alternative sources and procedures help to compare results and make decisions on them.

To assess the thematic accuracy classification map, confusion matrixes seem to be the most common way in pixel classification. However in object-oriented classification it is realizable to evaluate the results in terms of thematic classification and accuracy of the object edges.

In this study the STEP (acronym for Shape-Theme-Edge-Position) method was applied (Lizarazo, 2014). It allows evaluating the objects in the four terms, but for small objects. Since the study area

³ https://landsat.usgs.gov/sci_an.php (February, 2016).

and its covers are large, just partial results were obtained with STEP method. Therefore, additional research and testing must be done on this issue: an object-based classification is not complete until it has been assessed.

6.3. Training and skills

A valuable lesson learned is that to achieve the objective it was required:

- To operate as a team work. It includes remote sensing practitioners, statisticians and thematic experts, at least.
- To have favorable conditions provided by the institutional authorities.
- The awareness and clear exposition of ideas generated enthusiasm and commitment.
- A learning curve is being traced. Nothing has been definitely finished.
- Fluid communications are required to reach agreement on the decision-making process and its implementation.
- In addition to people, research and technology must be pillars of the process.

6.4. Cost-Benefit Analysis

In the project free images and software were used. Therefore, costs were significantly reduced. However, due to the lack of data in 2005 image (poor quality) and the poor software documentation, further work was needed to overcome these limitations.

New techniques and knowledge became available to DANE and can be used in the implementation of new projects.

It was evidenced that the free tools Interimage and R can be considered enough to meet the project requirements.

In general terms, results were satisfactory and showed that it is possible to overcome the challenges and to replicate this project for other areas.

6.5. Moving forward

With the project a way has been explored, but many others can be implemented too.

Big Data from satellite images is a source of relevant information for the calculation of some SDGs. According to the Global Survey on the use of BIG DATA, this source is not widely considered for the computation of SDGs indicators. This is an opportunity to encourage its global use in official statistics.

Benefits generated by the use of remote sensing, create the opportunity to obtain more traceable information and allow different types of multi-temporal studies, which gives a great added value for environmental and land use studies.

The methodology provided an overview of the object-based methods and demanded the combined use of GIS and image processing: spectral information and context for classification.

The operations that were implemented on commercial software can be also done on free tools as Grass, Ilwis, GV SIG or Qgis.

REFERENCES

- Atlántico, C. -C., CORMAGDALENA, & Internacional, D. -C. (2007). *Plan de ordenamiento y manejo de la cuenca hidrográfica del río Magdalena en el departamento del Atlántico*. Atlántico.
- Baatz, M., & Schape, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg*, 12-23.
- Battha. (2010). *Analysis of urban growth and sprawl from remote sensing data*. New York: Springer.
- Cambell. (2016). *Geological and Environmental Remote Sensing Laboratory*. Recuperado el 06 de 02 de 2016, de Principles of remote sensing:
http://gers.uprm.edu/geol3105/pdfs/05_remotesensing.pdf
- CORMAGDALENA; CRA; DAMB. (2006). *Plan de ordenamiento y manejo de la cuenca hidrográfica de la ciénaga de Malloquín*. Atlántico.
- Costa et al. (2013). Applying Multiresolution Segmentation Algorithm to Generate Crop Management Zones based on Interpolated Layers. *Sustainable Agriculture through ICT innovation*. Torino, Italia.
- DANE. (15 de 01 de 2016). *DANE*. Obtenido de <http://www.dane.gov.co/index.php/poblacion-y-demografia/proyecciones-de-poblacion>
- Frohn, R., Autrey, B., Lanes, C., & Reif, F. (2008). Segmentation and object-oriented classification of wetlands in a karst Florida landscape using multi-season Landsat-7 ETM+ imagery. *International Journal of Remote Sensing*.
- Gronemeyer, P. (30 de 10 de 2015). *The landscape Toolbox*. Obtenido de http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:object-based_classification
- Happ, P., Ferreira, R., Bentes, C., Costa, G., & Feitosa, R. (2010). Multiresolution segmentation: a parallel approach for high resolution image segmentation in multicore architectures. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38.
- Hay, & Castilla. (2006). Object- Based Image Analysis: Strengths, Weaknesses, Opportunities and Threats (SWOT). *International Archives of Photogrammetry*.
- Jensen. (2005). *Introductory Digital Image Processing: A Remote Sensing Perspective*. Pearson - Prentice-Hall.
- Lizarazo, I. (2014). Accuracy assessment of object-based image classification: another STEP. *International Journal of Remote Sensing*, 35, 6135-6156.
- R Foundation. (2 de 12 de 2015). *R Project for Statistical Computing*. Obtenido de <https://www.r-project.org/>

- Riggan, & Weih . (2010). *Object-based classification vs. pixel-based classification: Comparative importance of multi-resolution imagery*. (R. S. The International Archives of the Photogrammetry, Ed.) Recuperado el 04 de 02 de 2016, de http://www.isprs.org/proceedings/XXXVIII/4-C7/pdf/Weih_81.pdf
- Rossiter, D. (2014). Technical Note: Statistical Methods for Accuracy Assesment of Classified Thematic Maps. *Department of Earth Systems Analysis University of Twente, Faculty of Geo Information Science & Earth Observation (ITC), 25, 107.*
- Schowengerdt. (2007). *Remote sensing models and methods for image processing*. New York: Academic Press.
- SDSN. (20 de 01 de 2016). *Indicators report*. Obtenido de <http://indicators.report/indicators/i-68/>
- United Nations. (2015). *Sustainable Development Goals – The 2030 Agenda for Sustainable Development and the Sustainable Development Goals*.
- Xia, & Liu . (2010). Assessing object-based classification: advantages and limitations. *Remote Sensing Letters. Remote Sensing Letters, 187 – 194.*

7. DISCUSSION QUESTIONS

1. One of the great challenges of this project is replicate for major cities in Colombia, quickly and efficiently. But first we must overcome issues related to access to quality images, storage and processing of large volumes of data.

The question for the participants is: Do you know of any strategic alliance have been established between satellite information providers and public institutions to access satellite images at very low cost, for statistical purposes and public policy?

Regarding to storage and processing large volumes of information, we want to know if you have developed or known methodologies to facilitate automated processing or processing large numbers of images?

2. The data geovisualization techniques are an important method for communication and analysis of large volumes of information. What kind of scientific visualization tools can be used to disseminate the results of Big Data projects? Can the users interact with data in different ways using scientific visualization tools?