ANALYSIS OF THE RELATIONSHIP BETWEEN INTRA-URBAN VEGETATION CHANGE AND SOCIO-ECONOMIC DATA

L. M. G. Fonseca [a], G. A. Boggione [a,b], A. M. V. Monteiro [a], R. Santos [a]

[a] INPE, National Institute for Space Research, Brazil - (leila, giovanni, miguel)@dpi.inpe.br, rafael.santos@lac.inpe.br
[b] IFG, Federal Institute of Goiás, Brazil

Commission II, WG II/3

KEY WORDS: Data mining, Fusion, Vegetation, Urban, Change.

ABSTRACT:

Understanding the vegetation dynamics in urban areas in both quantitative and qualitative aspects is essential to population welfare and also to economic, social and environmental development. However, it is necessary appropriate tools for monitoring and analysis of the landscape dynamic in a systematic way. Therefore, this study proposes a methodology to analyze the relationship between intra-urban vegetation and the social-economic data using the integrated techniques of remote sensing and GIS as well as data mining. This research intends to answer questions such as: Is it possible to extract the intra-urban vegetation as well as identify the intra-urban vegetation changes from medium spatial resolution images and digital image processing techniques? Is it possible to establish a relationship between the intra-urban vegetation changes and social-economic information using data mining techniques?

1. INTRODUCTION

The lack of effective policies for giving order to the development of cities and their rapid growth are related, in most cases, to the many consequences of urbanization. The quest in understanding the diversity in the aspects of urban space, related to their physical-territorial dimensions and their inhabitants has become a concern for urban management and planning. In this sense, studies related to urban environmental quality (UEQ) has been increasingly frequent. Among all the variables used to evaluate UEQ, vegetation is recognized as the key one for many reasons: filtering air, water, and sunlight; cooling urban heat; recycling pollutants; moderating local urban climate; providing shelters to animals and recreational areas for people (Liang and Weng, 2011].

The integration of remote sensing technology and Geographic Information Systems (GIS) has been of paramount importance, since they allow the investigation of the landscape dynamic and the alleged correlation with social and economic variables.

Therefore, this research aims at proposing a methodology to identify the current conditions of the existing intra-urban vegetation. This study was carried out under the following assumptions:

- The intra-urban vegetation can be observed and quantified from orbital remote sensing images that are processed to be set to suitable conditions to the investigation on the basis of their spectral and spatial characteristics

- The effects of seasonal climate affect the concentration of water of the vegetation. This effect may be noticed in satellite images recorded in the dry and rainy seasons. If the vegetation coverage is receiving periodic care, these seasonal variations should be softer than those of vegetation that do not receive artificial care.

In this way, it becomes possible to better diagnose the situation of the city in relation to its intra-urban vegetation. With this diagnosis, spatial indicators that address the management condition of the vegetation coverage can be defined.

To carry on the analysis, a case study in the city of Goiânia, Brazil, between the years of 2008 and 2009 was performed. The methodological approach involves the data integration of

Remote Sensing, GIS and Data Mining techniques to generate a scenario that permits an exploratory analysis in the relation between intra-urban vegetation and the socioeconomic conditions in the city of Goiânia, Brazil. The digital image processing techniques are used to improve the visual quality of the images as well to highlight all the information of interest for the use of the human analyst, in turn leading to a range of applications.

2. METODOLOGY

The city of Goiânia (Figure 1) is located approximately 190 km from the Capital, between the coordinates 49O27' W, 16O50' S e 49O04' W, 16O27' S, as can be seen in Figure 1. It occupies a total area of 740.53 km2, of which about 40% of the city's area is already urbanized.

With a total resident population of 1,093,007 inhabitants (IBGE, 2000) the city of Goiânia suffered a significant population boom from the 1950s to the 1980s. At this time, the population of the city nearly doubled every ten years, probably due to the change of the capital to Brasília, and also due to the government projects of infrastructure and incentives for the use and occupation of the Cerrado biome for agricultural practice. After the 1980s, the population growth has remained high in Goiânia, with about 20% every 10 years.

The rapid population growth that occurred in Goiânia also induced the rapid growth of the urban area. What is most interesting is that although most of the city still belongs to the rural zone (60%), virtually the entire population of Goiânia (99.34%) of the total population resides in urban areas, and the rest of the population, only (0.66%) resides in rural areas.

The remote sensing images used in this study were downloaded through the Catalog provided by National Institute for Space Research (http://www.dgi.inpe.br), described below:

- Satellite image from LANDSAT-5 satellite, TM sensor, band 2 (green), 3 (red), and 4 (infrared), Point Orbits 222/071 and 222/072, corresponding to the Goiânia region; acquisition date on 03/03/2009, rainy season, spatial resolution of 30m.

- Satellite image from CBERS 2B, CCD sensor, bands 2 (green) 3 (red), and 4 (infrared), Point Orbit 158/119;

acquisition date on 08/19/2008, dry season, spatial resolution of 20m.

• Satellite image from CBERS 2B, HRC sensor, panchromatic band, and 158/A/119/2 Point Orbits; acquisition on 10/10/2008, dry season, spatial resolution of 2.5m.

For the image geo-referencing step, an ortophoto was used in 2006, and to validate the extraction of vegetation a QuickBird Image of 2002 was used. Both images have spatial resolution of 60 cm, and we use true color composition (no infrared).

Figure 1 – City of Goiânia, Brazil. Urban area appears in red color.

Census data, correspondent to the 2000 census, was obtained at the Brazilian Institute of Geography and Statistics (IBGE). This aggregated data was made available by the Census Districts.

Additional vector data was acquired together with the city of Goiânia in the form of MUBDG (Basic Urban Digital Map of Goiânia) and were updated in 2008.
 (http://www.goiania.go.gov.br/html/geoprocessamento/mapa. htm).

Figure 2 presents the flowchart of the methodology carried out for processing the satellite images, and then to generate the vegetation change map. This methodology was adapted from Domingos (2005). Following, we provide a detailed description about the digital image processing techniques used in this study.

2.1 Image Pre-processing

Due to differences in the images generated by the sensors, it is natural that distortions between the images occur. Thus some pre-processing steps are necessary to correct the data so that they become consistent to the proposed procedure.

The adjustment in question began with mosaic of images from TM and HRC sensors as shown in Figure 3. Next, we performed the atmospheric correction through the subtraction technique of the dark pixel (Chavez, 1988). Subsequently, we processed the images with a restoration filter, which improves the effective spatial resolution of the image and interpolates them at a finer sampling grid (Fonseca et al., 1993). The pixel size of CCD (20 m) and TM (30 m) was changed to 10 meters using the aforementioned restoration algorithm.

Finally, the data were interpolated (cubic convolution) to 2.5m to present the same pixel size of image HRC, a fundamental condition for the success of the fusion process.

For geo-referencing HRC, CCD and TM images, we used a 2006 ortophoto, in UTM Projection system with Datum Sad 69, as reference. For this, a set of 47 identifiable and well distributed control points throughout the study area were collected, having a pixel error of less than 0.38. To further

minimize the registration error, the REGEEMY (http://regima.dpi.inpe.br/) system, version 0.2.43, was used, which allowed a refinement of the control points with an error below 0.17 pixel.

As the scenes went beyond the area of interest after mosaicking, it was necessary to superimpose the image with a vector file with the limits of the region of interest, to eliminate the area that would not be used.
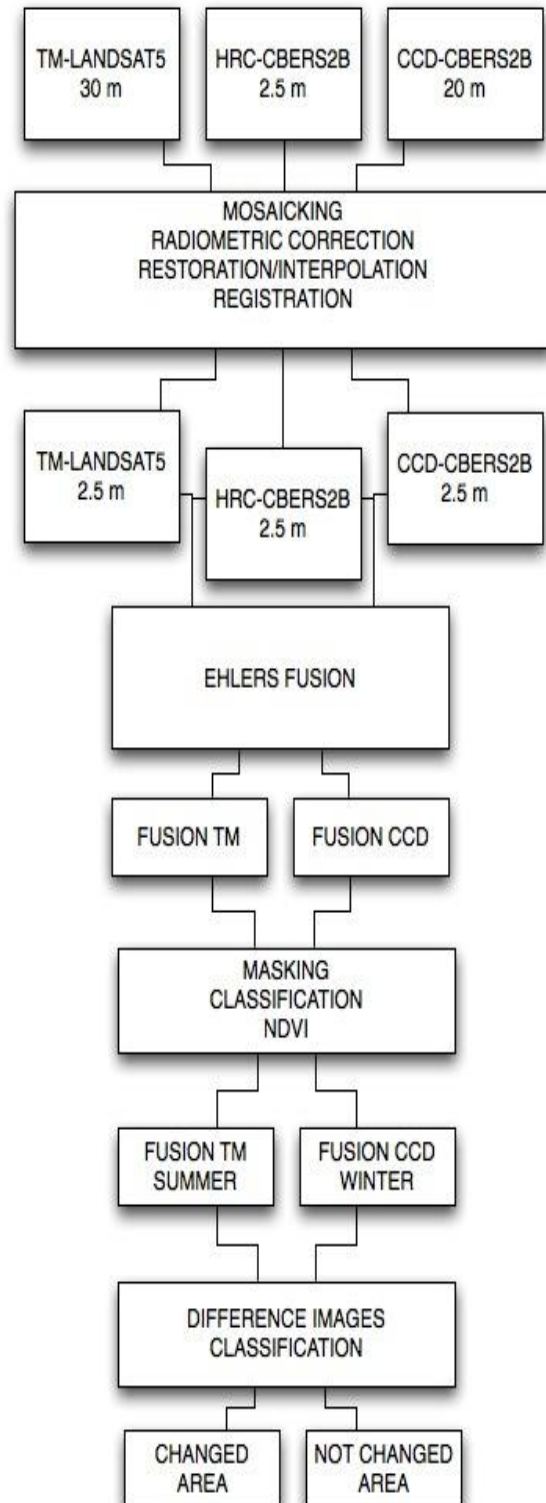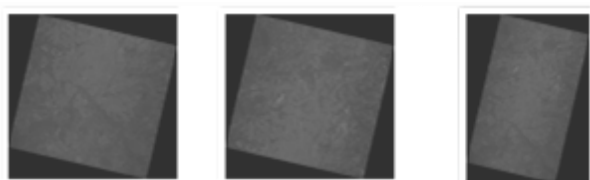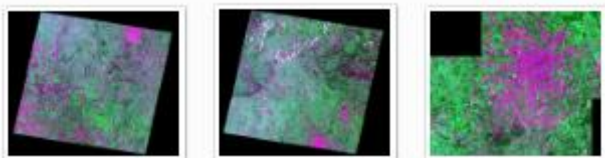

Figure 2 – Methodological Flowchart.

(a)   Path/Row: 158/A/119_1 and 158/A/119_2.

(b)   Path/Row: 222/071 and 222/072.

Figure 3 – Image Mosaicking (a) HRC and (b) TM.

We used a limit vector of the pilot area in Goiânia (Figure 4) that contains 5 macro-areas (Campinas, South, Central, Macambira and Vale do Maia Ponte) to produce a mask that delimits the area of interest, instead of using the whole scene, as shown in Figure 5. These areas were chosen due to their high population and large construction density.
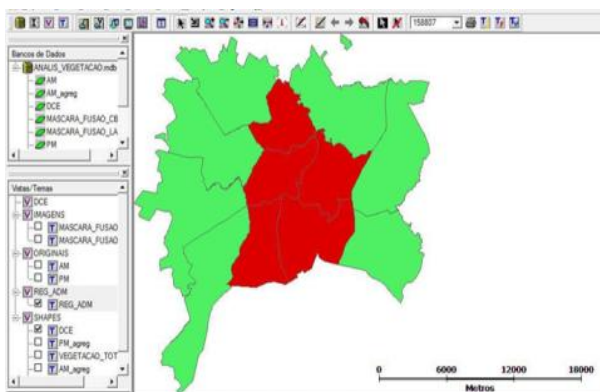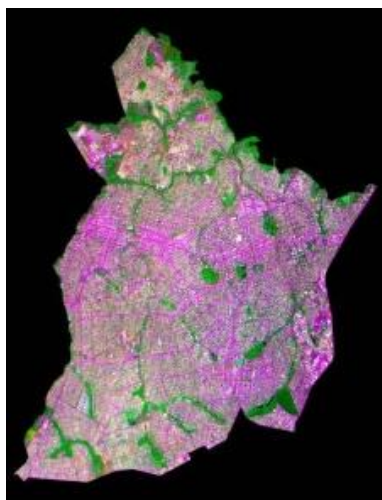

Figure 4 – Mask Generation (in red).

Figure 5 – Study Area.

2.2 Image and Data Processing

The fusion of HRC panchromatic band and multispectral images from TM and CCD sensors were carried out by using the EHLERS fusion method (Ehlers, et al., 2010). Figures 6 and 7 show details of the fused images, which have spatial resolution of 2.5 m.

The next step was to generate a mask that would identify only the vegetation areas in the hybrid images. We applied the HIS transformation to the hybrid TM image (3R4G2B) to generate the hue component (H), which was used to identify the vegetation in the image.

Figure 6 – TM images (3R4G2B), Goiânia Airport:
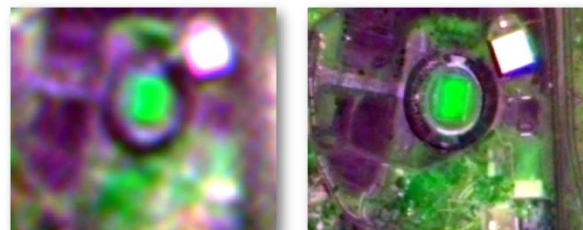Multispectral image (left) and hybrid image (right).

Figure 7 – CCD images (3R4G2B), Serra Dourada Stadium:
(left) and hybrid image (right).

At this stage only the hybrid TM image was used because, by the hypothesis initially raised, this image should contain more areas with high spectral response for vegetation for being in the rainy season. Figure 8 presents the Hue image which shows the predominant vegetation in brighter tones.
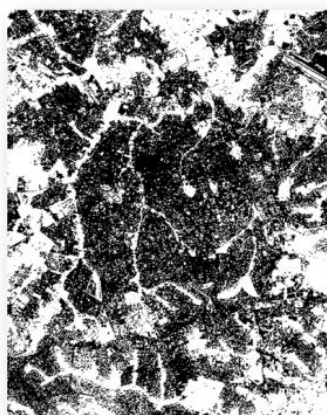

Figure 8 – Hue image for identifying the vegetation.

To identify and separate the vegetation targets from the non-vegetation targets, the Hue image was classified into two

classes: "vegetation" and "non-vegetation." For that, we used the Isoseg classifier, which is a non-supervised algorithm (Ball and Hall, 1965). It assumes no prior knowledge about the probability density distribution of the classes. From this resultant classification we generated a vegetation mask that identifies only vegetation targets.

Using the vegetation mask, we isolated the regions in the image that contain the intra-urban vegetation in the hybrid images. Afterwards, we generated NDVI (Normalized Difference of Vegetation Index) images for each season, RAINY NDVI and DRY NDVI (Schowengerdt, 1997). According to the premise of this study, the NDVI images present different aspects in relation to the greenness information. Therefore, the subtraction between them should present differences, which eventually might exist. The difference NDVI (NDVI_Dif) image is presented in Figure 9.
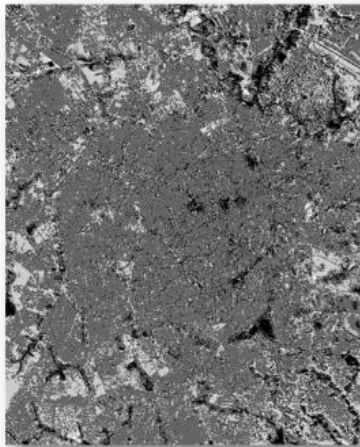

Figure 9 – NDVI_Dif image: difference between RAINY NDVI and DRY NDVI images.

Next, we classified the NDVI_Dif image using the supervised minimum distance classifier to distinguish two classes: Significant Change (higher values in the NDVI_Dif) and Small Change (lower values in the NDVI_Dif). The minimum distance classifier is indicated when the size of the training sets is small (Wacker and Landgrebe, 1972). The result obtained is shown in Figure 10, where green and red colours stand for, respectively, the Significant Change, and Small Change classes.
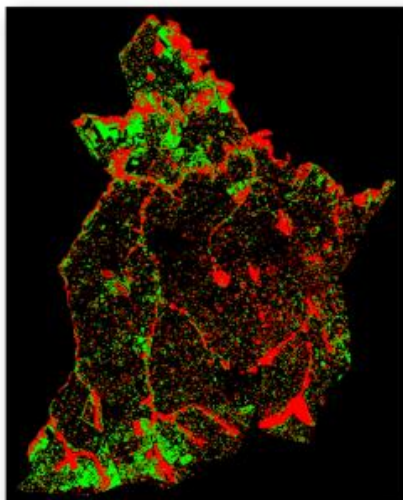

Figure 10 – Classification by the Minimum Distance method: Significant Change (green), and Small Change (red) .

Following, the classified image was vectorized. Each class generated a set of polygons with a unique identification. These vectors were exported in shape format (standard format of vector dada) as 2 files: Significant Change (SC) and Small Change (LC) as shown in Figure 11.
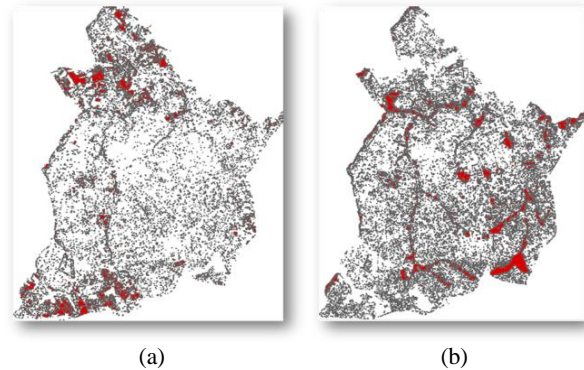

Figure 11 – vegetation change map: (a) Significant Change (SC), and (b) Small Change (LC).

Given the vegetation change map, the next step was to establish a relationship between vegetation changes and social-economic information. To accomplish this, firstly we obtain the intersection between the Areas of Interest and the Census Districts, in which the result is the area by census districts.

Besides, we obtain the Total vegetation (TV) by summing LC and SC polygon areas. By crossing the LC, SC, VT polygons and the Census Districts we obtain the vegetation changes by District (D) as illustrated in Figure 12. Considering that there might be large districts with small amount of vegetation or the opposite, we normalized the amount of vegetation per district area. This procedure is done to ensure that no district that has a small area and a lot of vegetation be placed equally with another that has the same quantity of vegetation in a larger area.
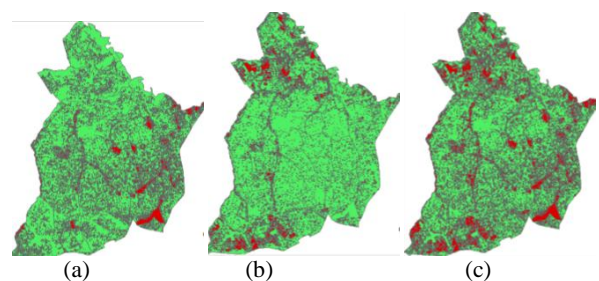

Figure 12 – Normalized LC, SC, and TV vegetation per District.

3. RESULTS

The relationship between vegetation changes and income is presented in Figure 13. We can observe that districts with high income have small vegetation changes. Differently, for districts with low income the vegetation changes are larger. Broadly speaking, the class "Small Change" is that one most present on the map, with about 52% of vegetation coverage, whereas the class "Significant Change" corresponds to 48% of the total vegetation coverage in the test area.

Figure 14 illustrates a vegetation area that received care. This region was classified in the image as "Small change" class. In a GIS framework we can enquiry information by attribute. Below, we give some examples of queries to explore the features of SIG.
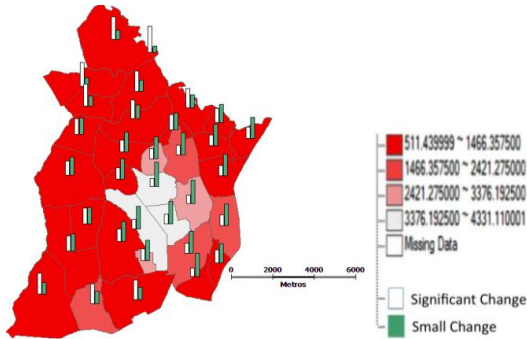

Figure 13 – Relationship between vegetation changes (LC and SC) and income.

Figure 14 – Vegetation area pointed out as "Small change" in the vegetation change map.

Figure 15 presents a spatial query involving the SC variable and income. We asked the system to indicate the districts that have vegetation change higher than 0.8 and income lower than R$ 1,244. The system pointed out the districts of Vale do Meia Ponte, which was confirmed by the field work as shown in Figure 16. This area presents abandoned flower beds covered with vegetation and city squares that are unattended.
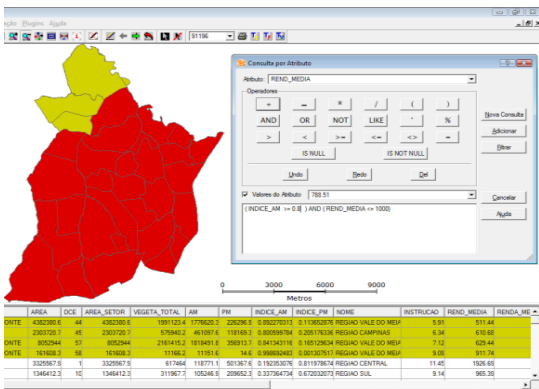

Figure 15. Spatial query: districts with amount of vegetation higher than 0.8 and income lower than R$ 1,244 (yellow).

Figure 16 –Significant Changes in the Vale do Rio Meia Ponte.

Conversely, we may ask what are the districts with vegetation change less than 0.2 and income higher than R$ 3.500,00 (Figure 17).
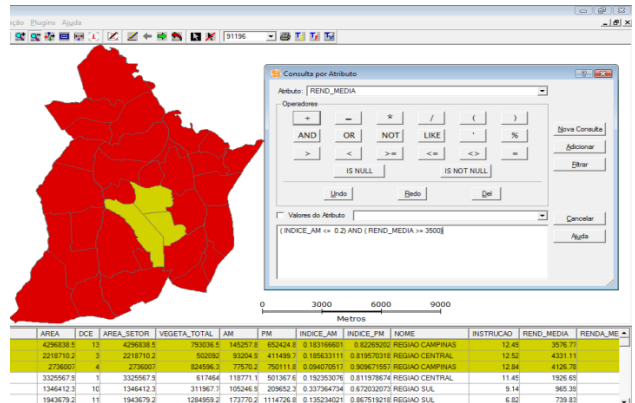

Figure 17. Spatial Query: what are the districts with vegetation change less than 0.2 and income higher than R$ 3.500,00.

The Central Sector and the Campinas Sector appeared as districts with better income and low change, which are also confirmed, as illustrated in Figure 18.

Figure 18 – Small Change in the Central Sector.

We also verified the possibility of establishing a relationship between intra-urban vegetation changes and social-economic

information using data mining algorithms to find association rules between the variables. Association rules allow the identification of items or facts that frequently occur together in a database (Tan, Steinbach and Kumar, 2009).

To run the association rule creation algorithm we first obtained socioeconomic data for the 40 districts in the city. This data complements the data about vegetation change in these districts. The data was processed so that a collection of regular square regions with 20x20 meters was created. Each of these regions contains the socioeconomic and vegetation change data for the corresponding geographic region. Finally, from this data we've selected only a subset where there are extreme changes in vegetation.

Association rules algorithms require that the data are represented with discrete values, which was done using arbitrary intervals that correspond roughly to the concepts low, high, etc. To extract the association rules we have used the Apriori algorithm implemented in the software WEKA (Waikato Environment for Knowledge Analysis) (Witten and Frank, 2000). This algorithm accepts as input a table of discrete entries and presents as output a set of association rules ordered in such a way that most frequent associations (list of discrete values that occur together in the database) appear first, allowing the manual evaluation of the results.

The most relevant results obtained with Apriori algorithm are presented below. These following results were extracted from a list of 300 most significant associations obtained with the algorithm:

- Where there is a high change of vegetation, there is also the presence of drainage;
- Where there is more education and higher income, there is small change;
- Where there is higher education and recycling, there is small change;
- Where there is lower education and high population density, there is higher change.

We can observe that the maintenance of vegetation has a relationship with the socioeconomic quality of the region. In areas where income and the level of education are lower, changes in vegetation are more frequent. The opposite scenario is also true. Areas with higher socioeconomic quality present lower intra-urban indices of vegetation change.

Other associations appear and are not as intuitive, such as the relationship between population density per household and low wages leading to a condition of high vegetation change.

4. CONCLUSION

The proposed methodology presented coherent results in relation to the in situ analysis. Regions with higher income families have more cared intra-urban vegetation. Suburban regions (lower income families), with worst conditions, presented less cared or even abandoned vegetation areas. Data mining techniques allowed the identification of not so clear relationships when using traditional techniques as GIS. This study also showed that SIG, digital image processing and data mining techniques are useful to extract information about environmental variables to subsidize public policies for the maintenance of public green areas. This study also confirms several hypotheses about the relationship between urban planning/growth and the environmental degradation which will in turn affects the life quality of Brazilian population.

5. REFERENCES

Chavez, P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sensing of Environment, 24, pp.459–479.

Domingos, P. L. H. Application of orbital remote sensing data of high resolution analysis of intra-urban vegetation and its conservation status. 2005. (INPE-TDI/). Master Thesis in Remote Sensing. National Institute for Space Research, São José dos Campos.

Fonseca, L. M. G., Mascarenhas, N. D. D., Prasad, G. S. S. D. Combined Interpolation-Restoration of Landsat Images Through a FIR Filter Design Techniques, 1993. IEEE Transactions on Geoscience and Remote Sensing, 14, pp.1247 - 2561.

Geoffrey H. Ball, David J. Hall, 1965. Isodata: a method of data analysis and pattern classification, Stanford Research Institute, Menelo Park, United States. Office of Naval Research. Information Sciences Branch.

Ian H. Witten and Eibe Frank, 2000. Data Mining - Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers.

IBGE, Brazilian Institute of Geography and Statistics. Data from the population census of Brazilian's Cities. http://www.ibge.gov.br (25 Jun. 2012).

Liang, B.; Weng, Q., 2011. Assessing urban environmental quality change of Indianapolis, United States, by the remote sensing and GIS integration. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sensing, 4, pp.43-55.

Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introdução ao Data Mining - Mineração de Dados, Editora Ciência Moderna, 1ª edição, 2009.

Schowengerdt, R.A. Remote Sensing: models and methods for imageprocessing. New York: Academic Press, 1997. 523p.

Wacker , A. G. and D. A. Landgrebe , Minimum Distance Classification in Remote Sensing, LARS Information Note 030772, Laboratory for Applications of Remote Sensing, Purdue Univers ity, W. Lafaytte , Indiana 47907, February, 1972.