

GeoDMA—Geographic Data Mining Analyst



Thales Sehn Körting*, Leila Maria Garcia Fonseca, Gilberto Câmara

Image Processing Division – DPI, Brazil's National Institute for Space Research, INPE, São José dos Campos, Brazil

ARTICLE INFO

Article history:

Received 27 July 2012

Received in revised form

2 January 2013

Accepted 20 February 2013

Available online 1 May 2013

Keywords:

OBIA

Image processing

Data mining

Image segmentation

Multitemporal analysis

Landscape ecology

ABSTRACT

Remote sensing images obtained by remote sensing are a key source of data for studying large-scale geographic areas. From 2013 onwards, a new generation of land remote sensing satellites from USA, China, Brazil, India and Europe will produce in 1 year as much data as 5 years of the Landsat-7 satellite. Thus, the research community needs new ways to analyze large data sets of remote sensing imagery. To address this need, this paper describes a toolbox for combing land remote sensing image analysis with data mining techniques. Data mining methods are being extensively used for statistical analysis, but up to now have had limited use in remote sensing image interpretation due to the lack of appropriate tools. The toolbox described in this paper is the Geographic Data Mining Analyst (GeoDMA). It has algorithms for segmentation, feature extraction, feature selection, classification, landscape metrics and multi-temporal methods for change detection and analysis. GeoDMA uses decision-tree strategies adapted for spatial data mining. It connects remotely sensed imagery with other geographic data types using access to local or remote database. GeoDMA has methods to assess the accuracy of simulation models, as well as tools for spatio-temporal analysis, including a visualization of time-series that helps users to find patterns in cyclic events. The software includes a new approach for analyzing spatio-temporal data based on polar coordinates transformation. This method creates a set of descriptive features that improves the classification accuracy of multi-temporal image databases. GeoDMA is tightly integrated with TerraView GIS, so its users have access to all traditional GIS features. To demonstrate GeoDMA, we show two case studies on land use and land cover change.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Remote sensing data is the only source that provides a continuous and consistent set of information about the Earth's land and oceans (Bradley et al., 2007). Combined with ecosystem models, remotely sensed data offers an unprecedented opportunity for predicting and understanding the behavior of the Earth's ecosystem (Tan et al., 2001b). Since 1970s, the Landsat series of satellites has provided optical images of the land's surface of the Earth every 16 days at a resolution of 30 m. The Landsat archive at the United States Geological Survey contains about 1 petabyte and is fully accessible worldwide (Câmara et al., 1996). From 2013 onwards, a new generation of optical remote sensing satellites from USA, China, Brazil, India and Europe will produce in 1 year as much data as 10 years of the current Landsat-7 satellite. Space agencies worldwide are operating or planning around 260 Earth observation satellites over the next 15 years. These satellites will carry over 400 different instruments, including optical and radar sensors for land imaging, gravity instruments, and ocean color

cameras. Our methods to analyze and understand massive data sets lag far behind our ability to produce and store this data (Dial et al., 2003; Fayyad et al., 1996; Wassenberg et al., 2009).

Working with large data sets of remote sensing data, researchers can produce results of large scientific and social impact (Bruzzone et al., 2003). Making effective use of these large data sets needs advances in GIScience (Goodchild, 2004). Remote sensing imagery provides information on land cover, which does not translate directly into land use and land change information (McCauley and Goetz, 2004). Therefore, to extract information about land change, we need to better represent the semantic content of remote sensing imagery (Câmara et al., 2001). In our view, the key to extracting land change information from remote sensing data is to develop methods that aim to capture landscape dynamics. Thus, the segmentation methods that are used to extract objects from the images have to be tuned not to find fixed objects, but to find regions that are subject to change in relation to the rest of the image (Smith, 1995). These regions are then mined by statistical methods that can capture landscape dynamics.

During 1980s and 1990s, most remote sensing image analysis techniques were based on per-pixel statistical algorithms (Blaschke, 2010). These techniques aimed at representing the knowledge about land cover patterns in terms of a limited set of parameters, such as average and standard deviation values of

* Corresponding author. Tel.: +55 12 3208 6522; fax: +55 12 3208 6468.

E-mail addresses: tkorting@dpi.inpe.br (T.S. Körting), leila@dpi.inpe.br (L.M. Garcia Fonseca), gilberto@dpi.inpe.br (G. Câmara).

groups of individual pixels. Recently, Object-Based Image Analysis (OBIA) has shown to be a good alternative to traditional per-pixel and region based approaches. Differently, OBIA approaches first identify regions in the image, extract neighborhood, spectral and spatial descriptive features and afterwards combine regions and features for object classification. Although segmentation has a large tradition in image processing (Haralick and Shapiro, 1985) and remote sensing (Câmara et al., 1996), OBIA took a long time to reach mainstream users. This approach became popular when it combined image segmentation with good labeling methods that match the features to those of user-defined classes. Most successful software packages, either proprietary like eCognition (Lang et al., 2007) or ENVI Feature Extraction (ITT, 2008), or the open sources InterIMAGE (Costa et al., 2010) and geoAIDA (Bückner et al., 2001), make use of semantic networks in the analysis process. Semantic networks contain prior knowledge about the specific characteristics of object classes and their interrelations. However, remote sensing image analysis using OBIA can be lengthy and complex because of the processing difficulties related to image segmentation, the large number of features to be resolved (Pinho et al., 2008) and the many different methods needed to model the semantic networks (Hay and Castilla, 2008).

Despite the success of using semantic networks in the image analysis, one important challenge is the feature selection phase. We have to find metrics that best describe the region properties as well as select features that best distinguish between regions. Current software can extract a huge amount of statistics (mean value, standard deviation), spatial (area, perimeter, shape), color, texture, and topology features (distance to neighbors, relative border). To obtain an accurate classification, the feature selection often relies on *ad hoc* decisions about what should describe an object. Another problem is that land cover classes in most environments are not pure, or spectrally homogeneous. To approach this problem, scene models for classification usually present a nested structure, analyzing scenes in multiple scales (Woodcock and Strahler, 1987). One way is to use an approach that makes some hypothesis about the object properties defined within an application context. Such theory would provide metrics to extract object properties. Within this context, landscape ecology can help to define metrics by elaborating landscape types as *ecologically meaningful units*. Such land units can be used as the basis for analysis and assessment (Groom et al., 2006).

Another concern is how to build a semantic network for the interpretation task. User experience shows that there are no simple rules for building such networks, and this task may require considerable time and expertise (Lang, 2008). On the other hand, the number of available features makes a detailed

feature exploratory analysis a time-consuming task and dependent on expertise. In this case, data mining techniques can be useful to extract information from large databases where objects being classified are described through many features. Examples of works that used data mining with remote sensing data include Tan et al. (2001a), Silva et al. (2008), Stein et al. (2009) and Pinho et al. (2012).

In spite of the considerable advances made over the last few years in high resolution satellite data, image analysis tools, and services, end users still lack effective and operational tools to help them manage and transform remote sensing data into useful information that can be used for decision making and policy planning purposes. For instance, Table 1 provides a summary of the software used in image analysis studies. It is observed that most works used more than one computational program to perform the analysis. This introduces more challenges for the researchers such as data integration, conversion of data format, knowledge of the software to be used, files replication, and other problems that make the data analysis process difficult. Consequently, the need of a framework capable of merging all image analysis tasks (such as segmentation, feature extraction and selection, data mining, pattern recognition and multi-temporal analysis) into a single platform is seen as a great demand. Although there are some good proprietary image analysis software available, the licensing costs can be a barrier for their use. Besides, these systems cannot be studied and adapted for ones own needs. Steiniger and Hay (2009) discussed all these problems in a review about the use of geographic information tools in landscape ecology, which are critical for any application. They also advocated that sharing knowledge through the development of Free and Open Source Software (FOSS) is a requirement for technological and scientific advancement.

Considering the aforementioned challenges, the contribution of this work is two-fold. Firstly we proposed and implemented a new toolbox, developed under the FOSS foundation, for integrating remote sensing imagery analysis methods with a repertoire of data mining techniques producing a user-centered, extensible, rich computational environment for information extraction and knowledge discovery over large geographic databases. The new toolbox is called GeoDMA—Geographic Data Mining Analyst. It integrates techniques of segmentation, feature extraction, feature selection, landscape and multi-temporal features and data mining, allowing pattern recognition tasks and multi-temporal analysis in large geographic databases. Secondly we developed an approach for multi-temporal analysis that allowed creating a new set of features based on polar coordinates transformation to describe temporal cyclic events such as those common in agriculture applications.

Table 1
Software used in the remote sensing applications.

| Article | ArcGIS | eCognition | ENVI | Fragstats | R | Weka | Others | Total |
|--------------------------------|--------|------------|------|-----------|---|------|------------|-------|
| Addink et al. (2007) | | × | | | × | | CAN-EYE | 3 |
| Esquerdo et al. (2009) | | | × | | | | | 1 |
| Ferraz et al. (2005) | × | | | × | | | SWAT2000 | 3 |
| Frohn and Hao (2006) | × | | × | × | | | | 3 |
| Gavlak et al. (2011) | | | | | | | GeoDMA | 1 |
| Hüttich et al. (2009) | | × | | | × | | TIMESAT | 3 |
| Imbernon and Branthomme (2001) | | | | × | | | ERDAS | 2 |
| Lackner and Conway (2008) | | × | | | | | | 1 |
| Lewinski and Bochenek (2008) | | × | | | | | PCI | 2 |
| Metzger et al. (2009) | × | | | | × | | | 2 |
| Novack et al. (2011) | | × | | | | × | InterIMAGE | 3 |
| Pinho et al. (2012) | | × | × | | | × | | 3 |
| Ribeiro et al. (2009) | × | | | | × | | ERDAS | 3 |
| Saito et al. (2011) | | | | | | | GeoDMA | 1 |
| Silva et al. (2008) | | | | × | | × | Spring | 3 |
| Southworth et al. (2002) | × | | | × | | | | 2 |

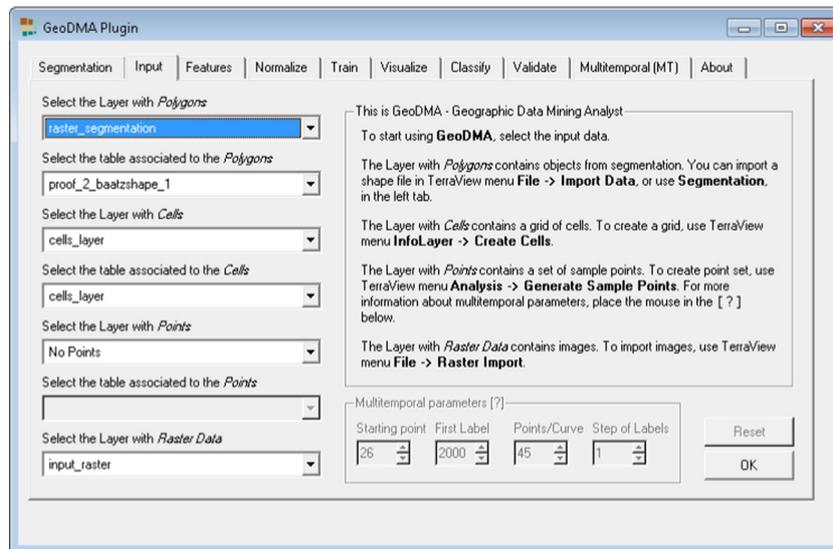


Fig. 1. User interface for GeoDMA.

In particular, GeoDMA was thought to provide some technical capabilities, which fulfill critical requirements (Steiniger and Hay, 2009) for geographic information tools in remote sensing applications. Below, we list the principal functionalities of GeoDMA:

1. support for different geographic data types in a local or remote database;
2. spatio-temporal analysis tools, including a visualization scheme for temporal profiles;
3. a set of features based on polar coordinates that allows describing temporal cyclic events as well as improving the classification accuracy of multi-temporal data;
4. simulations to assess the accuracy of process models (e.g. using Monte Carlo methods);
5. rapid creation of thematic maps and other results due to its integration on top of TerraView GIS (INPE, 2012b);
6. detection of multi-temporal changes as well as creation of change maps, allowing to explore the causes, processes and consequences of land use and land cover change (Saito et al., 2010).

In Section 2 we present the underlying methods used in GeoDMA, describing input data, segmentation of multispectral imagery, cycles detection in multi-temporal imagery, feature extraction and classification methods. Following, in Section 3 we provide 2 case studies with different target applications, exploiting the wide range of available features. In Section 4 we conclude and discuss future works.

2. GeoDMA description

GeoDMA is a system for image analysis which integrates image analysis tools, metrics based on landscape ecology theory, multi-temporal features handling, and data mining techniques (Körtling et al., 2008). The system is based on the methodology proposed by Silva et al. (2005), to identify deforestation patterns in the Amazon. It is a free software solution for remote sensing applications, running on different platforms, e.g. Windows and Linux. All processing modules are integrated in a Graphic User Interface (GUI), shown in Fig. 1.

The system works as a *plugin* to the software TerraView (INPE, 2012b), which provides the interface to the user (hereby called the interpreter), with visualization of geographic information data stored in databases. GeoDMA is coded in C++, using the QT

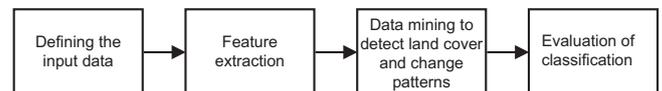


Fig. 2. GeoDMA: diagram of the main processing steps for image analysis.

cross-platform application development framework (Blanchette and Summerfield, 2008), and the free GIS and image processing library TerraLib (Câmara et al., 2008).

Fig. 2 shows a general diagram of the system. The processing modules start by defining the input data, going through feature extraction and the application of data mining algorithms to extract and deliver information about Earth observation. We describe each module of the GeoDMA system in the following.

2.1. Module “input data definition”

GeoDMA deals with a variety of geospatial data, stored in databases as raster¹ or vector formats. Object-based approaches use homogeneous regions from image segmentation. The regions extracted by the segmentation operation, points (pixels), and cells (treated as regular grids) define regions in vector format. Multi-temporal images can be represented as a sequence of raster snapshots, which are used to extract a sequence of values for each region in different intervals that define a curve called cycle.

2.1.1. Module “segmentation”

Image segmentation is one of the most challenging tasks in digital image processing. One simple definition states that a good segmentation should partition the image into regions with homogeneous behavior (Haralick and Shapiro, 1985). The system provides 4 segmentation algorithms:

- Region growing approach based on Bins et al. (1996). This algorithm defines random seeds over the image and merges them with neighboring pixels, according to a similarity threshold. According to Meinel and Neubert (2004), this algorithm produced results with good overall impression with proper delineation of homogeneous areas.
- Segmentation based on Baatz et al. (2000), a region growing and multi-resolution procedure. The interpreter defines the

¹ Throughout the text, image and raster terms will be used interchangeably.

parameters for scale, band's and color's weights, and region's weights for smoothness and compactness.

- A *chessboard* segmentation, which creates a set of square regions.
- An algorithm based on Korting et al. (2011b), which classifies spectrally similar pixels according to their location in the feature space, using a geographic extension of the Self-Organizing Maps (SOM).
- A technique of resegmentation applied to urban images based on Korting et al. (2011a). Resegmentation performs adjustments in a previous segmentation in which the elements are small regions with a high degree of spectral similarity (oversegmentation).

2.1.2. Cycles detection

Analysts interpret the imagery and map changes by analyzing differences found in images taken at different times. However, it is a tedious and time-consuming task to interpret long series using manual methods (Boulila et al., 2011). Studies to identify cyclic events have used images and products from the Moderate Resolution Imaging Spectroradiometer (MODIS), which is an important source of Earth data with high temporal resolution and low spatial resolution (Verbesselt et al., 2010). This imagery records photosynthetic activity, allowing the surface analysis in time and space (Jiang et al., 2008), and also provides vegetation index values (EVI2) in a spatial resolution of 250 m (Huete et al., 2002).

By following the EVI2 values in a certain position along the time, we can define a temporal profile that has a cyclic behavior, as seen in Fig. 3. This profile represents EVI2 values from 2000 to 2011, in a spatial resolution of 250 m pixel and temporal resolution of 16 days. The cyclic behavior should not be considered change, even though they contain different states as the variation from 0.15 to 0.85 between 2007 and 2008. However, techniques must be able to distinguish cycles to classify land cover and land change patterns. For this task, we can use temporal profiles to describe transitions between objects, and this way monitoring the land cover change dynamics (Freitas and Shimabukuro, 2008).

The land pattern detection in GeoDMA using multi-temporal images is based on cycles. Therefore it is important to distinguish the terms *profile* and *cycle*, although they represent the same temporal entity in some cases. Suppose we have a profile with observational data over a 5-year period while the analysis is performed yearly. In this case, the profile contains the full time series, divided into 5 cycles of 1 year. In vegetation analysis the cycles are often annual. For example, a time series with temporal resolution of 8 days define a cycle with around 45 values per year ($\frac{365}{8} \approx 45$). For 16 days there are 23 values for one cycle, and so on.

The central question is how to describe each cycle. Before using data to quantify or infer spatio-temporal processes, it is crucial to understand how the processes are represented in the data. Characterization of multi-temporal imagery provides insights into how different processes are represented by the spatial, spectral and temporal sampling of the imagery (Small, 2011). In agriculture applications the duration of certain events is well defined, e.g. 1 year. From multi-temporal images, the user defines the initial point of a cycle and the number of points for each cycle. With this information, GeoDMA is able to extract multi-temporal features from time series.

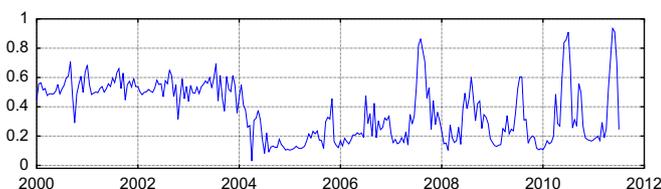


Fig. 3. EVI2 profile example from 2000 to 2011, the range of values is [0,1]. Adapted from Freitas et al. (2011).

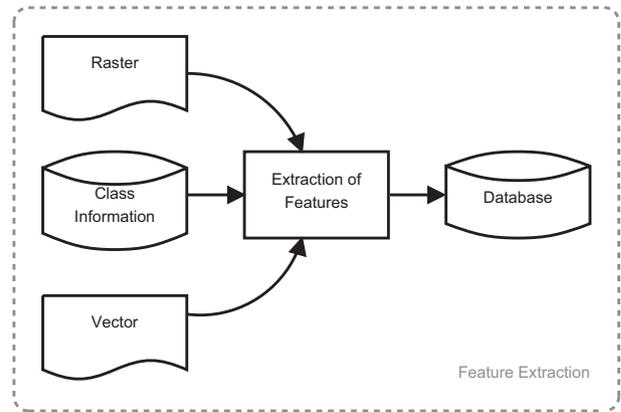


Fig. 4. Feature extraction—Spectral and Spatial features use raster and vector information. Landscape ecology features use class information. Multi-temporal features use cycles' information.

2.2. Module “feature extraction”

Fig. 4 describes the feature extraction module, which stores all extracted features in a local or remote database. According to the raster size and the quantity of regions this task can take long time to be performed. Therefore, the creation of a feature database ensures that all features will be extracted only once.

Features are divided into 3 groups. The *segmentation-based* features are properties obtained from the segmented regions, integrating raster and vector data types. The *landscape-based* features obtained from the landscape ecology metrics are stored as vector data. Cycles from raster time series are used to extract *multi-temporal* features.

2.2.1. Segmentation-based features

The segmentation-based features include spectral (Table 2) and spatial (Table 3) metrics to describe each region stored in the database. The spectral features relate all pixel values inside a region, therefore include metrics for maximum and minimum pixel values, mean values, or texture properties; some of the spectral features are based on Pacifici et al. (2009). The spatial features measure the shapes of the regions, including height, width, or rotation; some of them are based on McGarigal and Marks (1994). Fig. 5 shows the visual representation of both features.

2.2.2. Landscape-based features

Because of time and space discontinuities, the real world environments are patchy (Wiens, 1976), defining a landscape as a spatially heterogeneous area (Turner, 2005). The landscape ecology concepts employed in GeoDMA are the base to analyze the structure of the landscape, defining geometric and spatial metrics for the regions present in the landscape, viewed as a *mosaic* of elements aggregated to form the pattern of *patches*, *corridors* and *matrices* on land (Forman, 1995).

Landscape ecology mainly considers *patches* as areas, or categories, containing habitat, and the main focus is on conservation. However, to adapt these concepts to remote sensing, *patches* are also related to different types, such as a deforestation area in a forest region, or a region containing a roof in a urban imagery (El-Shaarawi and Piegorsch, 2002). Based on these considerations, 3 groups of metrics are defined (McGarigal, 2002):

- *Patch metrics* qualify individual patches and characterize their spatial and contextual information. Examples include the area of a polygon, perimeter, and compacity. As an example, one patch can be defined as a forest fragment.

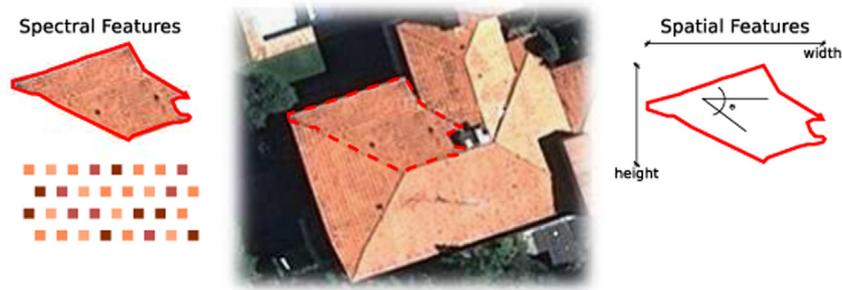


Fig. 5. Visual representation of the segmentation-based spectral and spatial features. Several features can be extracted from the highlighted region. Spectral features include metrics for maximum and minimum pixel values, or mean values. Spatial features measure the height, width, or rotation.

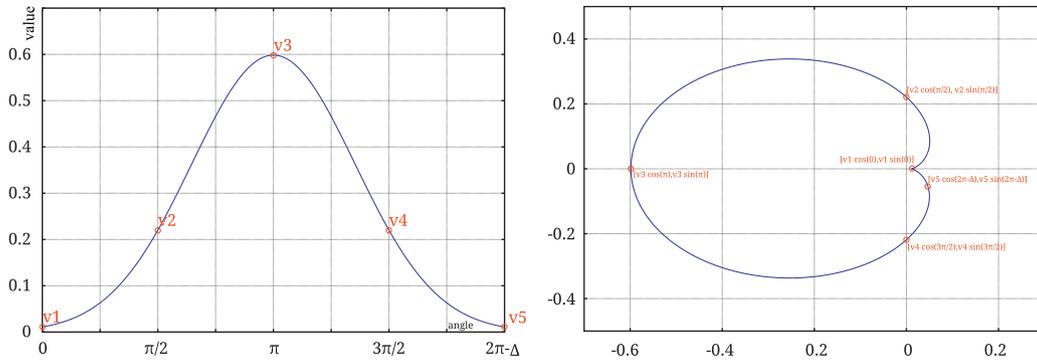


Fig. 6. When the values of the cycle are associated to a certain angle (left), the closed shape is created from its polar transformation (right).

- *Class metrics* integrate all patches of a given type inside a specific area, by simple or weighted averaging. The weighted averaging scheme can reflect a greater contribution of large patches to the overall index. Instances include average shape index, and patch size standard deviation. These metrics are used to define, for instance, the amount of houses in a block, or the average size of croplands in a state.
- *Landscape metrics* concern all patch types or classes inside a specific area. These metrics are integrated by a simple or weighted averaging, and they reflect combined patch mosaic. Landscape metrics include average perimeter–area ratio and patch size coefficient of variation.

Table 4 describes some types of landscape ecology features.

2.2.3. Multi-temporal features

Multi-temporal features include several descriptors for cycles. This group encompasses phenological indicators, a well known set of metrics for time series, as described in Pettorelli et al. (2005) and Hüttich et al. (2009), including the dates of the beginning or end of a growing season, the length of the green season, and so on. Besides *phenological*, we suggested to use the *linearity* metrics (Stojmenovic et al., 2008) and shape measures based on polar representation of cycles.

According to Hornsby et al. (1999), the standard computational models of time do not consider that certain events or phenomena may be recurring. The term cycle can also be used to capture the notion of recurring events. To support cycle's visualization, Edsall et al. (1997) proposed a time-wheel legend, resembling a clock face, divided into several wedges according to the data instances.

In our case, we adapted the time wheel legend by plotting each cycle of the profile, and by projecting values to angles in the interval $[0, 2\pi]$. Let a cycle be a function $f(x, y, T)$, where (x, y) is the spatial position of a point, and T is a time interval t_1, \dots, t_N , and N is the number of observations in such a cycle. The cycle can be

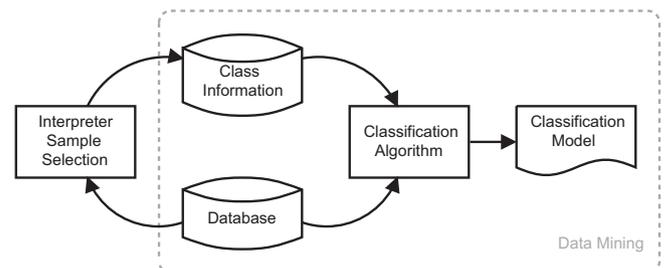


Fig. 7. The interpreter defines a typology and the set of representative samples, which are used to create the classification by applying the data mining algorithm.

visualized by a set of values $v_i \in V$, where v_i is a possible value of $f(x, y)$ in time t_i . Let its polar representation be defined by a function $g(V) \rightarrow \{A, O\}$ (A corresponds to the abscissa axis in the Cartesian coordinates, and O to the ordinate axis) where

$$a_i = v_i \cos\left(\frac{2\pi i}{N}\right) \in A, \quad i = 1, \dots, N \quad (1)$$

and

$$o_i = v_i \sin\left(\frac{2\pi i}{N}\right) \in O, \quad i = 1, \dots, N. \quad (2)$$

Considering $a_{N+1} = a_1$ and $o_{N+1} = o_1$, we can obtain the coordinates of a closed shape. Fig. 6 illustrates a cycle and its transformation to the polar coordinates. Given the shapes, we can extract various geometric features, such as area, perimeter, direction, or bounding ellipse. In this scheme, a cycle with constant values outcomes a circle, and different cycles draw different shapes according to their properties. Henceforth, this type of feature is named as *Polar*.

Moreover, a polar representation provides a new visualization scheme that can help us to describe the pattern represented in the cycle. A first idea when using annual cycles suggests splitting the

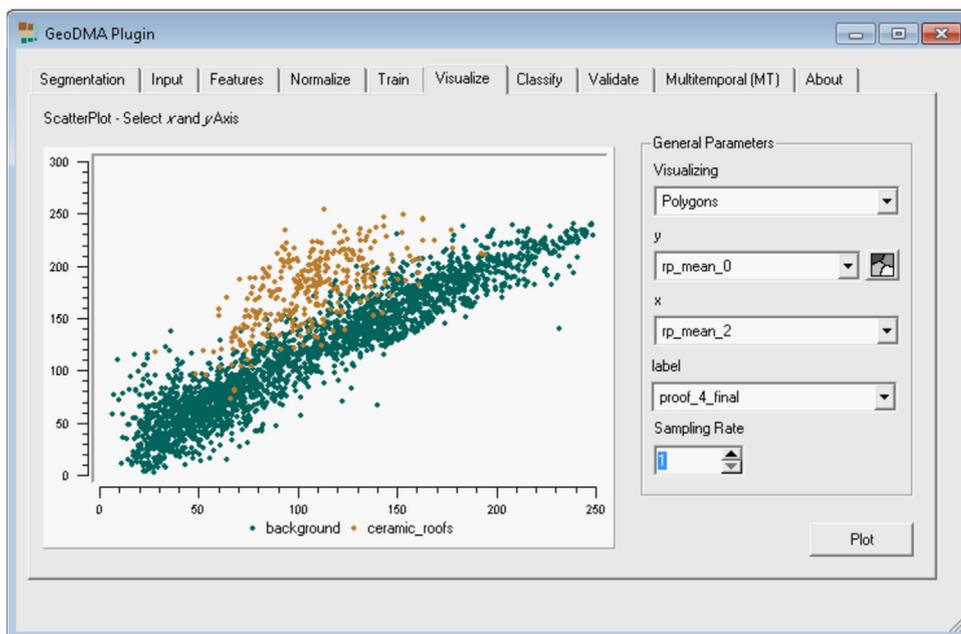
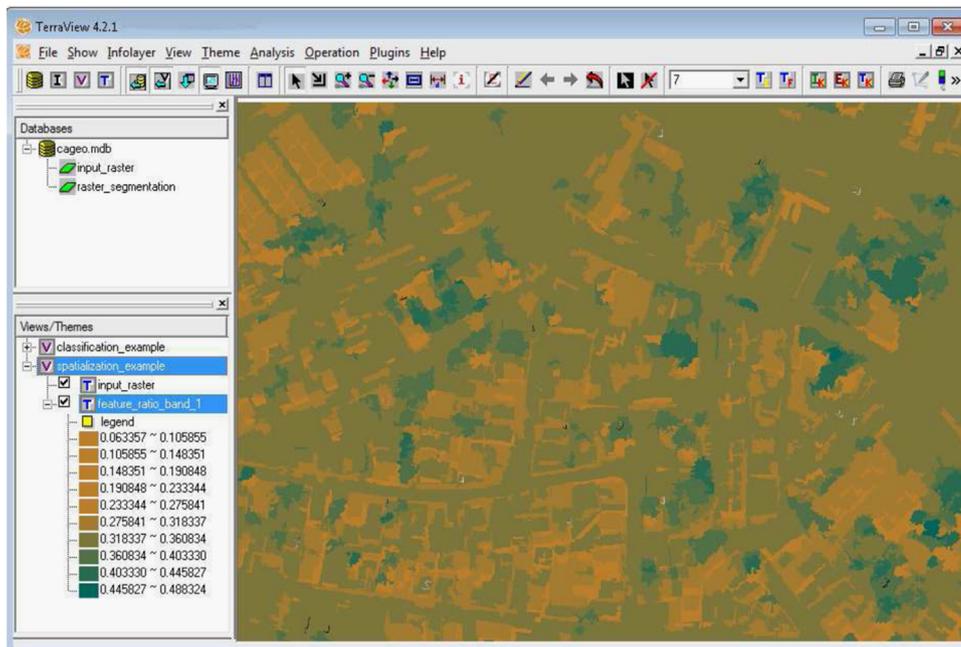


Fig. 8. The input image (top), visualization tools using a map of features (middle), and a scatterplot (bottom). The map shows the feature ratio of band 1, showing it is a proper feature to distinguish the target *Trees*. The scatterplot shows the feature space of features mean of bands 0 and 2, distinguishing the target *Roofs* from the rest of the image.

polar representation into 4 quadrants related to the 4 seasons. Hence, other features such as average values per season can also be computed. Another feature, called *Polar Balance*, calculates the standard deviation of the area per season, which indicates the stability of the profile throughout the cycles.

Phenological indicators plus linearity metrics are the commonly used metrics in the literature, therefore we refer to them as *basic* features. The remaining features are of the *polar* type. Table 5 describes the available multi-temporal features in GeoDMA.

2.3. Module “data mining for detecting land cover and change patterns”

In the data mining module (Fig. 7) the interpreter selects representative (training) samples of the expected patterns. All patterns compose the land cover typology, and some algorithm will create automatically a classification model based on training samples. The classification model can be stored for further and manual analysis. This model shall be used to classify the entire database, or different databases with the same expected typology.

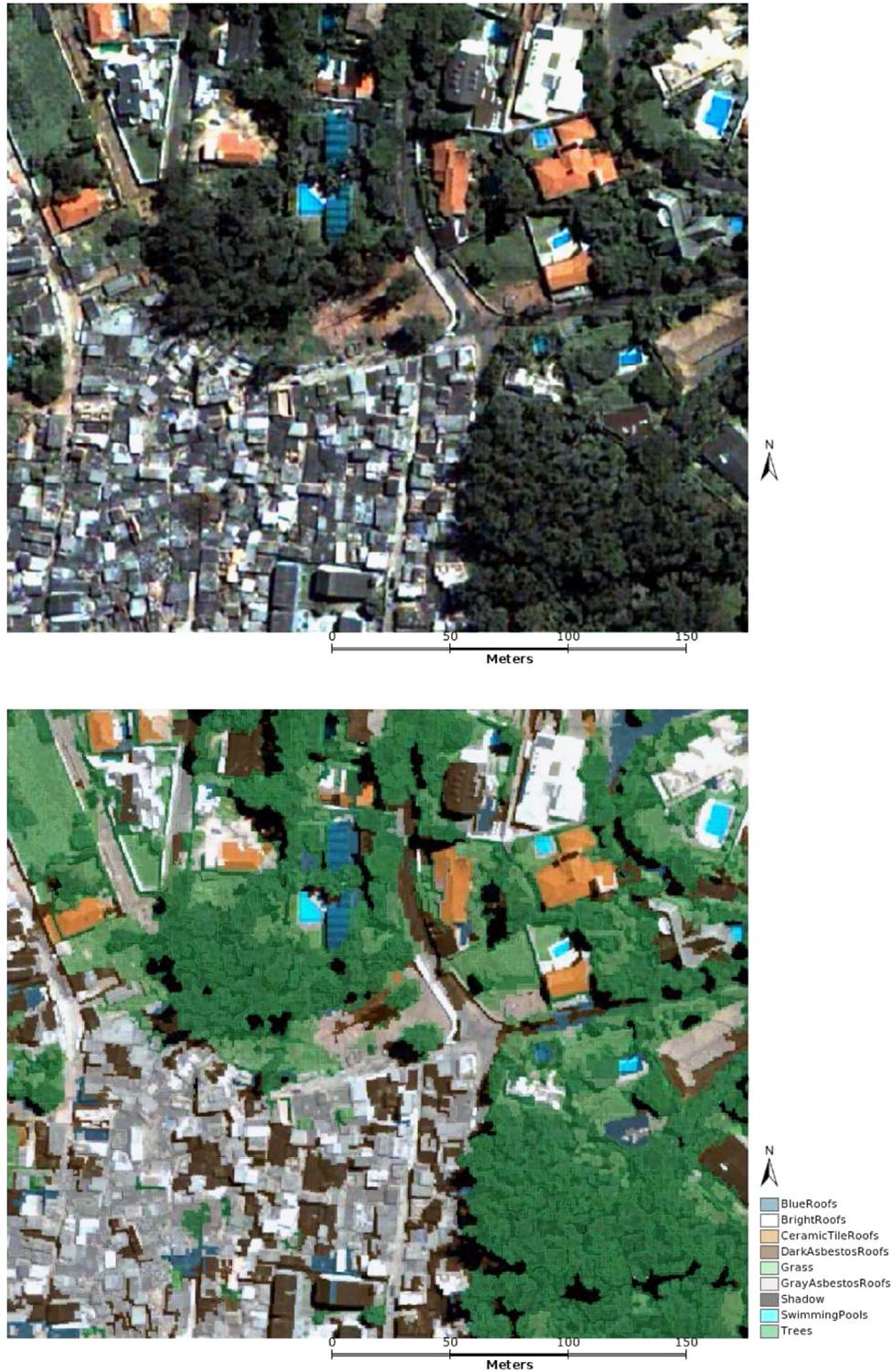


Fig. 9. Top: QuickBird image of São Paulo, southeast of Brazil, acquired on March 30, 2002, color composition R4G3B2. Bottom: intra-urban land cover map using GeoDMA.

Usually, the search for patterns includes the automatic execution of a classification algorithm and a phase of feature evaluation by the interpreter. As [Pinho et al. \(2012\)](#) point out, the inclusion of data mining techniques in the classification process can increase the speed and also reduces the empirical nature of the feature selection process and the creation of classification models. One mechanism to evaluate the features is provided in the visualization module, which displays features in a scatterplot to visualize the data distribution in the feature space, as shown in [Fig. 8](#).

According to [Fayyad and Stolorz \(1997\)](#), data mining is one step of a process called knowledge discovery in databases—KDD. In this sense, KDD involves data preparation, search for patterns, knowledge evaluation, and refinement, possible in multiple iterations. The search for patterns includes the automatic execution of a classification algorithm and the evaluation of features by the interpreter.

Since the interpreter knows the typology behind the data he must create a description of the expected patterns by selecting training samples for each pattern, which must be representative over the images. These samples are represented by a set of features. Afterwards, in the supervised classification step, the algorithm uses these training samples to build a classification model. Although GeoDMA provides 3 classification algorithms (decision trees, SOM, and neural networks) the focus of the experiments in this paper is on decision trees ([Quinlan, 1993](#)). In a classifier based on decision trees, thresholds are applied to object's features. Observations satisfying the thresholds are assigned to the left branch, otherwise to the right branch ([Hastie et al., 2009](#)). In the final step, classes are assigned to the terminal nodes (or leaves) of the tree.

2.4. Module "Classification evaluation"

The output of GeoDMA is a thematic map, created by applying the classification model to the database. According to [Fayyad and Stolorz \(1997\)](#), this step is part of a repetitive process, in which the interpreter evaluates the results visually and statistically. Depending on the obtained accuracy, the interpreter repeats some previous steps aiming to create a better classification model.

According to [Gamanya et al. \(2007\)](#), a strong and experienced evaluator of segmentation techniques is the human eye/brain combination. In addition, when dealing with multi-temporal analysis, the validation is often not straightforward, since independent reference sources must be available during the change interval ([Verbesselt et al., 2010](#)). However, it is always important to establish measures of correctness of the results with ground truth data. According to [Congalton \(2005\)](#), validation has become a standard component of any land cover map derived from remotely sensed data.

GeoDMA computes error matrices and the Kappa statistics ([Foody, 2002](#)) for a classification result. In the sample selection module the system automatically divides the samples into training and validation sets, randomly. Results are compared with validation samples to create the error matrix. In cases where the sample set is small, GeoDMA provides error evaluation based on Monte Carlo simulation ([Rubinstein and Kroese, 2008](#)) using only training samples.

3. Experimental results

In this Section we present 2 case studies to illustrate the effective use of the GeoDMA system. The first experiment uses segmentation-based features to map urban land cover classes. The second one uses multi-temporal features to map land cover classes.

3.1. Land cover classification of an intra-urban scene using high-resolution images

Identifying changes in land cover and land use provides important information for urban planning and management ([Meinel et al., 2001](#)). For example, this type of information can be used to plan changes to the public transportation system in areas in which the number of high-rise buildings is rapidly increasing. Such changes can be assessed using multi-temporal analyses of intra-urban land use and land cover maps, which require continuously updated, detailed and precise data ([Pinho et al., 2012](#)).

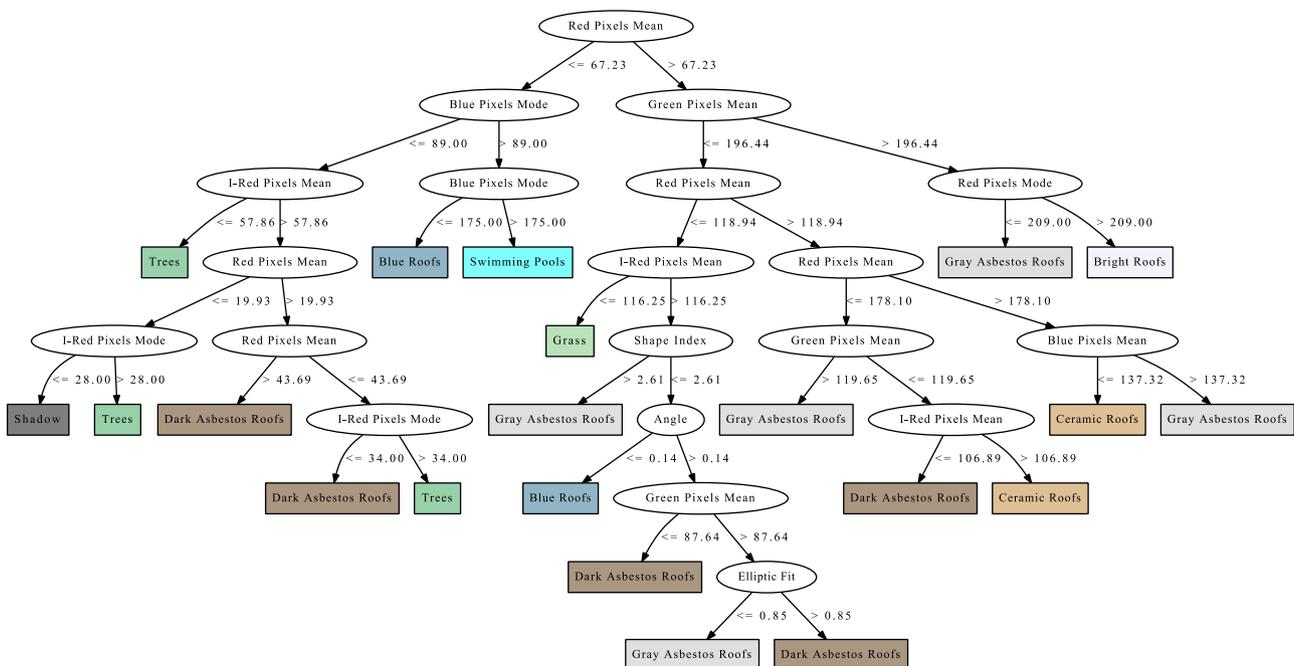


Fig. 10. The decision tree model for intra-urban land cover classification in a region of São Paulo city, Brazil.

To evaluate the effectiveness of GeoDMA system for land cover classification we conduct the study for the city of São Paulo, southeast of Brazil, with a great variety of intra-urban land cover classes, using QuickBird imagery. The images used in this experiment were acquired on March 30, 2002 and consist in a crop (523 × 445) of a hybrid multi-spectral image (0.6 m) with 4 bands blue, green, red, and infrared (Fig. 9, top).

The class typology includes roofs (blue, bright, ceramic, dark and gray asbestos), grass, swimming pools, shadows, and trees. The segmentation algorithm employed in this experiment is the multi-resolution procedure based on Baatz et al. (2000). The segmentation process created 2437 regions, and their corresponding geometrical and spectral features were extracted. In the training step, the interpreter labeled samples according to the

previously defined typology, with 15 training samples per class, and 10 validation samples per class.

All objects in the image were classified according to the model, and Fig. 9 shows the resultant thematic map. The classification model based on a decision tree was built using the previously selected samples. This model is illustrated in Fig. 10. The features used in this model included spectral mean values of the 4 bands, the mode values of blue, red, and infrared bands, besides the angle, shape index, and elliptic fit from the regions.

The land cover map was evaluated by Kappa coefficient, whose value was 0.84, with and overall accuracy of about 85%. Besides, the overall computational time to run GeoDMA was around 2 h, including the phases of feature extraction and sample selection by the interpreter.

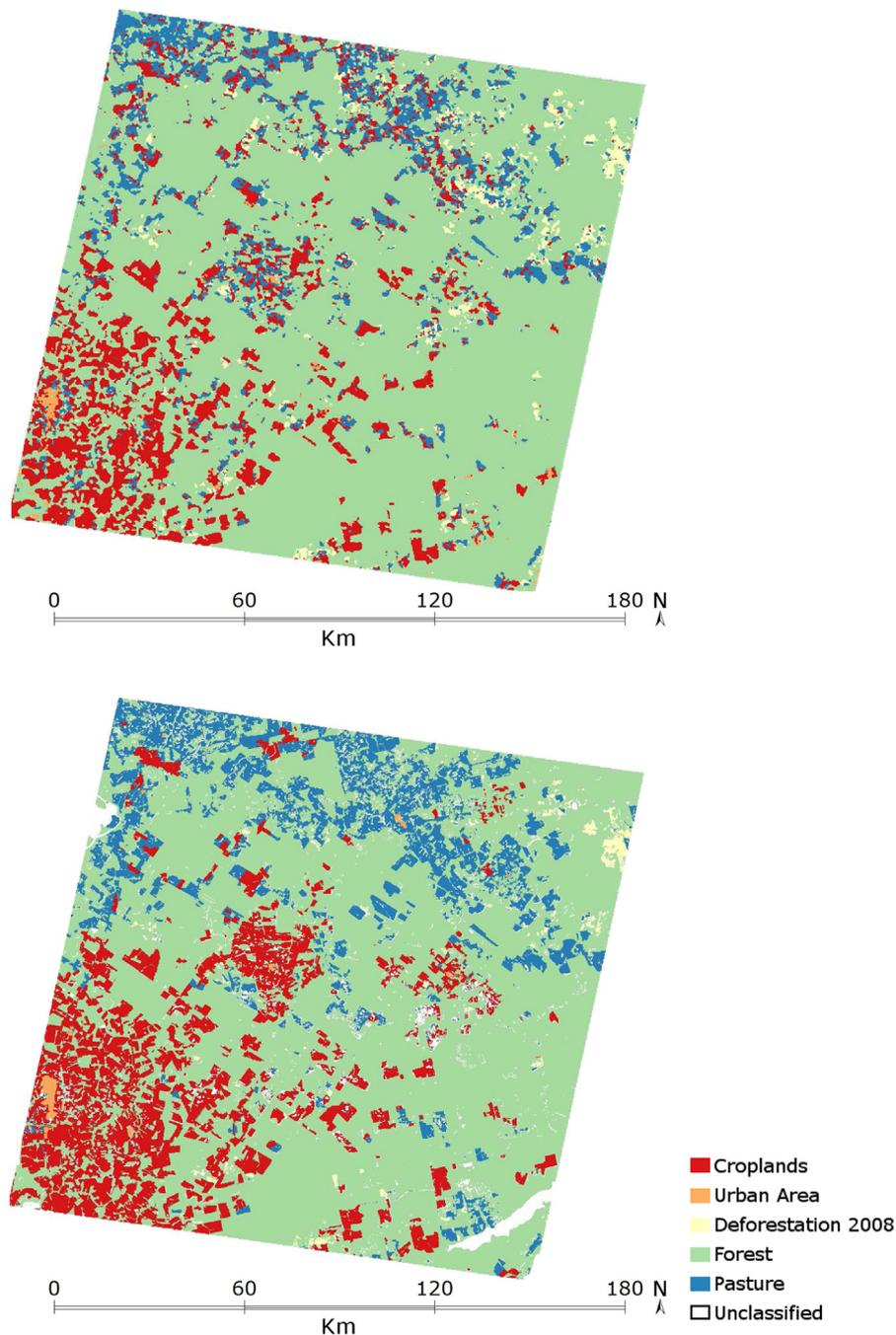


Fig. 11. Resultant land cover map for the second experiment (top), and the reference data for visual comparison (bottom).

3.2. Classification of multi-temporal imagery

In this experiment we employ GeoDMA to discriminate 5 land cover types in a Brazilian Amazon region. We used as reference the thematic maps produced by the project TerraClass (Almeida et al., 2009). It provides detailed land cover maps in deforested areas of the Brazilian Amazon for 2008. The deforested areas are estimated by the deforestation monitoring project named PRODES (INPE, 2012a). The typology includes *Croplands*, *Pasture*, *Urban Area*, *Deforestation 2008*, and *Forest*.

The study area included 14 088 samples randomly selected by the system (4000 samples for *Croplands*, 4000 for *Pasture*, 828 for *Urban Area*, 1260 for *Deforestation 2008*, and 4000 for *Forest*), located in the North of Mato Grosso (Lat. 11°d34'23"S, Lon. 54°43'14"W). Each sample represents a multi-temporal pixel corresponding to a cycle of 1 year. We selected cycles of vegetation indices from 2008 and the corresponding land cover from the reference data. The data sets used in these experiments consist of 16-day EVI2 profiles from MODIS with a 250 m pixel size, which is a Level 3 product (MOD13Q1), calculated from the Level 2 daily surface reflectance product (MOD09 series) (Vermote et al., 2002).

According to Verbesselt et al. (2010), validating multi-temporal land cover and land change methods is often not straightforward, since independent reference sources for a broad range of potential changes must be available during the change interval. In Fig. 11 we show the resultant classification using our model and the reference data for visual comparison.

The classification model resulted in a decision tree with 13 leaves, as shown in Fig. 12, and the accuracy resulted in a Kappa value of 0.82. By analyzing the model, one can observe the use of the basic features Area and Maximum values of the cycles. Besides the Sum and the Mean for the 1st slope of the cycles were used. Furthermore,

the polar features were Areas of the 1st and 2nd seasons, and the Polar Balance, which measures the variation of areas between seasons. The node of the tree which divides classes *Forest* and *Deforestation 2008* uses the feature Polar Balance with threshold of 0.14. A short variation between seasons describes a constant EVI cycle, which is expected in the class *Forest*. Instability in the EVI cycle produces higher values for this feature, and this fact is expected in cycle changes that occur in the class *Deforestation 2008*.

4. Concluding remarks

Remote sensing imagery provides information on land cover, which does not translate exactly into land use information (McCauley and Goetz, 2004). To produce valuable information about land, there exist several steps that if supported by computational tools deliver results in short time. In this sense, geographic data mining offers a cost-effective and fast alternative to deliver ancillary information that helps to understand the Earth and to predict further behaviors (Openshaw, 1999; Han and Kamber, 2008).

Therefore, we have developed the GeoDMA system, a free software that integrates image analysis tools for supporting different geographic data types in local or remote database, spatio-temporal analysis, and also a new set of feature descriptors based on polar coordinates that allows improving the classification accuracy. It provides an extensible set of features extracted from the scene objects, which can be represented as points, regions or cells. These features feed an automatic classification algorithm to model the discovered classes in one or more images. The system is used via a GUI including tools for visualization, typology definition, feature selection, classification, and evaluation. The multi-temporal

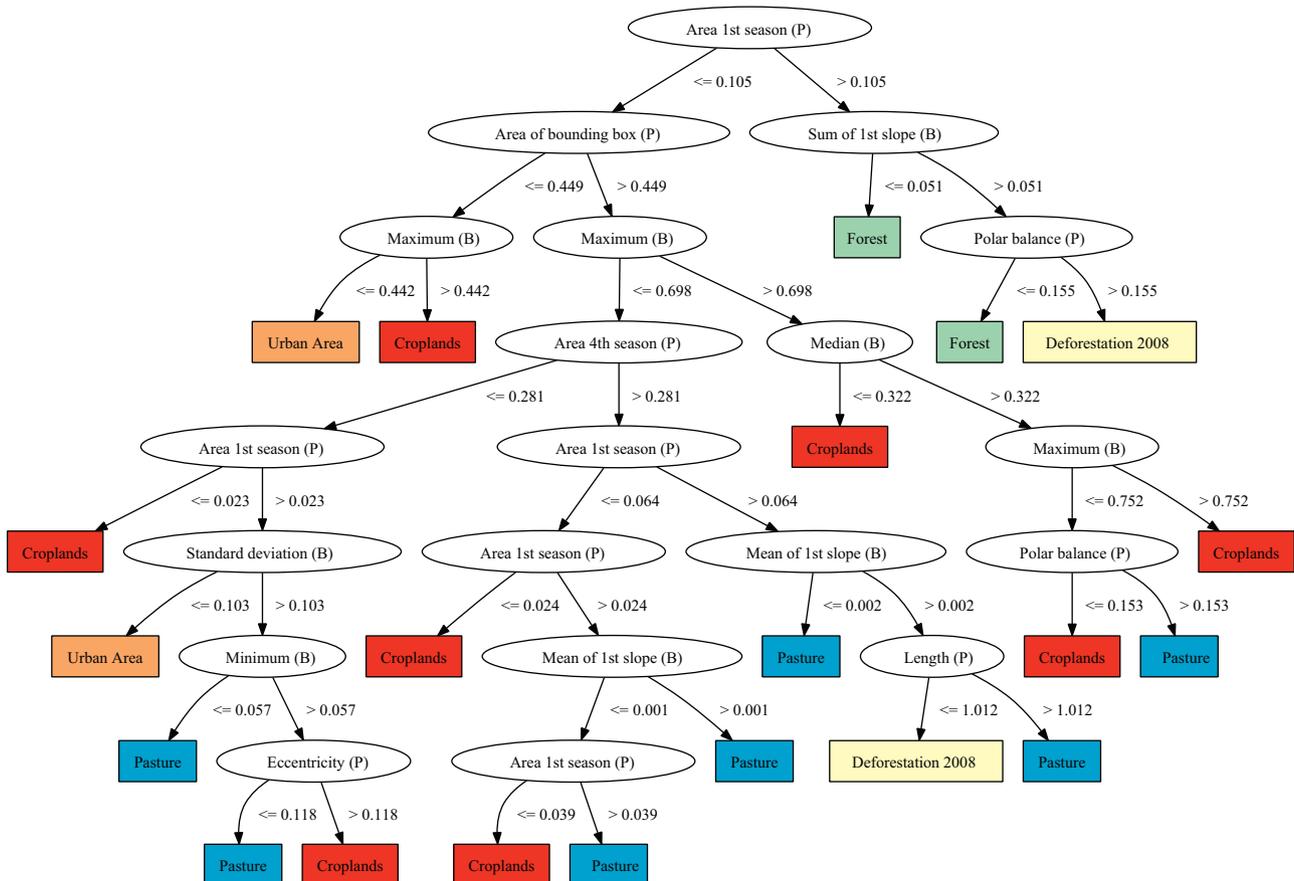


Fig. 12. Classification model for the second experiment, with 13 leaves and Kappa=0.82.

Table 2
Segmentation-based spectral features.

| Name | Description | Formula | Range | Unit |
|---------------|--|--|----------|------|
| Amplitude | Defines the maximum pixel value minus the minimum pixel value | $p_{X_{\max}} - p_{X_{\min}}$ | ≥ 0 | px |
| Dissimilarity | Measures how different are the GLCM elements. Higher values mean regions with high contrast | $\sum_{i=1}^{D-1} \sum_{j=1}^{D-1} p_{ij} \cdot i-j $ | ≥ 0 | – |
| Entropy | Measures the disorder in an image. When the image is not uniform, many GLCM elements have small values, resulting in large entropy | $-\sum_{i=1}^{D-1} \sum_{j=1}^{D-1} p_{ij} \cdot \log p_{ij}$ | ≥ 0 | – |
| Homogeneity | Assumes higher values for smaller differences in the GLCM | $\sum_{i=1}^{D-1} \sum_{j=1}^{D-1} \frac{p_{ij}}{1 + (i-j)^2}$ | ≥ 0 | – |
| Mean | Returns the average value for all N pixels inside the region | $\frac{\sum_{i=1}^N px_i}{N}$ | ≥ 0 | px |
| Mode | Returns the most occurring value (mode) for all N pixels inside the region. The first mode is assumption for multimodal cases | | ≥ 0 | px |
| Std | Returns the standard deviation of all N pixels (μ is the mean value) | $\sqrt{\frac{1}{N-1} \sum_{i=1}^N (px_i - \mu)^2}$ | ≥ 0 | px |

Table 3
Segmentation-based spatial features. The unit px means the amount of pixels.

| Name | Description | Formula | Range | Unit |
|----------------------|--|--|------------|-----------|
| Angle | Represents the main direction of a region. It is retrieved by the angle of the biggest radius of the minimum circumscribing ellipse | | $[0, \pi]$ | rad |
| Area | Returns the area of the region. When measured in pixels is equal to N | | ≥ 0 | px^2 |
| Box area | Returns the bounding box area of a region, measured in pixels | | ≥ 0 | px^2 |
| Circle | Relates the areas of the region and the smallest circumscribing circle. R stands for maximum distance between the centroid and all vertices | $1 - \frac{N}{\pi R^2}$ | $[0, 1]$ | px |
| Elliptic fit | Finds the minimum circumscribing ellipse to the region and returns the ratio between the area and the ellipse area | | $[0, 1]$ | – |
| Fractal dimension | Returns the fractal dimension of a region | $2 \frac{\log \text{perimeter}/A}{\log N}$ | $[1, 2]$ | – |
| Gyration radius | Equals the average distance between each pixel position in one region and its centroid. Smaller values stand for regions similar to a circle | $\frac{\sum_{i=1}^N pos_i - pos_c }{N}$ | ≥ 0 | px |
| Length | It is the height of the region's bounding box | | ≥ 0 | px |
| Perimeter | It is the amount of pixels in the region's border | | ≥ 0 | px |
| Perimeter area ratio | Calculates the ratio between the perimeter and the area of a region | $\frac{\text{perimeter}}{N}$ | ≥ 0 | px^{-1} |
| Rectangular fit | It is the ration between the region's are and the minimum rectangle outside the region. Higher values stand for regions similar to a rectangle | | $[0, 1]$ | – |
| Width | It is the width of the region's bounding box | | ≥ 0 | px |

Table 4
Landscape-based features. When the unit is hectares, the value is divided by 10^4 .

| Name | Description | Formula | Range | Unit |
|-----------------------|--|---|----------|---------|
| Class area | The metric CA means the sum of areas of a cell | $\sum_{j=1}^n a_j$ | ≥ 0 | ha |
| Percent land | $\%Land$ equals the sum of the areas (m^2) of all patches of the corresponding patch type, divided by total landscape area (m^2). $\%Land$ equals the percentage the landscape comprised of the corresponding patch type | $\frac{\sum_{j=1}^n a_j}{A} \times 100$ | $[0, 1]$ | % |
| Patch density | PD equals the number of patches of the corresponding patch type divided by total landscape area | $\frac{n}{A}$ | ≥ 0 | Patches |
| Mean patch size | MPS equals the sum of the areas (m^2) of all patches of the corresponding patch type, divided by the number of patches of the same type | $\frac{\sum_{j=1}^n a_j}{n} \cdot 10^{-4}$ | ≥ 0 | ha |
| Patch size std | $PSSD$ is the root mean squared error (deviation from the mean) in patch size. This is the population standard deviation, not the sample standard deviation | $\sqrt{\frac{\sum_{j=1}^n (a_j - MPS)^2}{n}} \cdot 10^{-4}$ | ≥ 0 | ha |
| Landscape shape index | LSI equals the sum of the landscape boundary and all edge segments (m) within the boundary. This sum involves the corresponding patch type (including borders), divided by the square root of the total landscape area (m^2) | $\frac{\sum_{j=1}^n e_j}{2\sqrt{\pi \times A}}$ | ≥ 1 | – |

module of GeoDMA currently deals only with pixel-based applications. Further work in this area includes the development of segmentation techniques for multi-temporal images and the integration of regions to obtain multi-temporal features.

Two case studies were presented to show the potential of GeoDMA in analyzing land patterns. The first study suggested that coupling spectral and geometric features with sample selection is a powerful strategy to quickly classify urban remote sensing images. The second demonstrated the classification of multi-temporal imagery with coarse spatial resolution images.

Further research is needed to explore the automatic detection of trajectories in time series, as well as the evaluation of trends for land cover change based on past events. According to Silva et al. (2008), patterns found in one map can be linked to those in earlier and later maps, enabling a description of the objects' trajectory of change. The polar features provide a novel way to describe multi-temporal profiles, and therefore can be investigated to map trajectories of changes in such data. Considering the land cover classification using imagery acquired for a single date, the development of GeoDMA needs improvements in some aspects. One of them is to extend the

Table 5
Multi-temporal features for describing cyclic events.

| Name | Type | Description | Range |
|---------------------|-------|---|---------|
| Amplitude | Basic | The difference between the cycle's maximum and minimum values. A small amplitude means a stable cycle | [0, 1] |
| Area | Polar | Area of the closed shape. A higher value indicates a cycle with high EVI values | ≥0 |
| Area per Season | Polar | Partial area of the closed shape, proportional to a specific quadrant of the polar representation. High value in the summer season can be related to the phenological development of a cropland | ≥0 |
| Circle | Polar | Returns values close to 1 when the shape is more similar to a circle. In the polar visualization, a circle means a constant feature | [0, 1] |
| Cycle's maximum | Basic | Relates the overall productivity and biomass, but it is sensitive to false highs and noise | [0, 1] |
| Cycle's mean | Basic | Average value of the curve along one cycle | [0, 1] |
| Cycle's minimum | Basic | Minimum value of the curve along one cycle | [0, 1] |
| Cycle's std | Basic | Standard deviation of the cycle's values | ≥0 |
| Cycle's sum | Basic | When using vegetation indices, the sum of values over a cycle means the annual production of vegetation | ≥0 |
| Eccentricity | Basic | Return values close to 0 if the shape is a circle and 1 if the shape is similar to a line | [0,1] |
| First slope maximum | Basic | It indicates when the cycle presents some abrupt change in the curve. The slope between two values relates the fastness of the greening up or the senescence phases | [-1, 1] |
| Gyration radius | Polar | Equals the average distance between each point inside the shape and the shape's centroid. Smaller values stand for shapes similar to a circle | ≥0 |
| Polar balance | Polar | The standard deviation of the areas per season, considering the 4 seasons. Small value point to constant cycles, e.g. the EVI of water (with a small Area), or forest (with a medium Area) | ≥0 |

architecture of the system to implement multiple scales analysis. When relating objects using multiple scales, new features can become available, such as the hierarchical relations.

Acknowledgments

The authors would like to thank CNPq, CAPES, and FAPESP (Brazil) for providing financial support to develop this research, and the anonymous reviewers for their valuable comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.cageo.2013.02.007>.

References

- Addink, E., Jong, S., Pebesma, E., 2007. The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. *Photogrammetric Engineering & Remote Sensing* 73 (8), 905–912.
- Almeida, C., Pinheiro, T., Barbosa, A., Abreu, M., Lobo, F., Silva, M., Gomes, A., Sadeck, L., Medeiros, L., Neves, M., Silva, L., Tamasauskas, P., 2009. Metodologia para mapeamento de vegetação secundária na Amazônia Legal. Technical Report, Brazil's National Institute for Space Research, São José dos Campos. URL: (<http://www.inpe.br/cra/>).
- Baatz, M., Schape, A., Schäpe, M., 2000. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. In: Wichmann-Verlag (Ed.), *XII Angewandte Geographische Informationsverarbeitung*. Herbert Wichmann Verlag, Heidelberg, pp. 12–23.
- Bins, L., Fonseca, L., Erthal, G., Li, F., 1996. Satellite imagery segmentation: a region growing approach. *Simpósio Brasileiro de Sensoriamento Remoto* 8, 677–680.
- Blanchette, J., Summerfield, M., 2008. C++ GUI Programming with Qt 4. Prentice Hall. URL: (<http://portal.acm.org/citation.cfm?id=1406186>).
- Blaschke, T., 2010. Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (January (1)), 2–16.
- Boullila, W., Farah, I., Ettaba, K., Solaiman, B., Ghézala, H., 2011. A data mining based approach to predict spatiotemporal changes in satellite images. *International Journal of Applied Earth Observation and Geoinformation* 13 (June (3)), 386–395.
- Bradley, B., Jacob, R., Hermance, J., Mustard, J., 2007. A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment* 106 (January (2)), 137–145.
- Bruzzone, L., Smits, P., Tilton, J., 2003. Foreword special issue on analysis of multitemporal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing* 41 (November (11)), 2419–2422.
- Bückner, J., Pahl, M., Stahlhut, O., Liedtke, C., 2001. geoAIDA—a knowledge based automatic image data analyser for remote sensing data. In: *Congress on Computational Intelligence Methods and Applications, CIMA*. ICSC, Bangor, Wales, UK.
- Câmara, G., Egenhofer, M., Fonseca, F., Monteiro, A., 2001. What's in an image? In: *Spatial Information Theory. Lecture Notes in Computer Science*, vol. 2205, pp. 474–488.
- Câmara, G., Souza, R., Freitas, U., Garrido, J., Li, F., 1996. Spring: integrating remote sensing and Gis by object-oriented data modelling. *Computers and Graphics* 20 (May (3)), 395–403.
- Câmara, G., Vinhas, L., Ferreira, K., Queiroz, G., Souza, R., Monteiro, A., Carvalho, M., Casanova, M., Freitas, U., 2008. TerraLib: an open source GIS library for large-scale environmental and socio-economic applications. In: *Open Source Approaches in Spatial Data Handling 2 (Advances in Geographic Information Science)*, pp. 247–270.
- Congalton, R., 2005. Thematic and positional accuracy assessment of digital remotely sensed data. In: *Proceedings of 7th Annual Forest Inventory and Analysis Symposium*. USDA Forest Service, pp. 149–154.
- Costa, G., Feitosa, R., Fonseca, L., Oliveira, D., Ferreira, R., Castejon, E., 2010. Knowledge-based interpretation of remote sensing data with the interimage system: major characteristics and recent developments. In: Addink, E., Van Coillie, F. (Eds.), *GEOBIA. ISPRS Working Groups*, Gent, Belgium. URL: (<http://geobia.ugent.be/>).
- Dial, G., Bowen, H., Gerlach, F., Grodecki, J., Oleszczuk, R., 2003. IKONOS satellite, imagery, and products. *Remote Sensing of Environment* 88 (November (1–2)), 23–36.
- Edsall, R., Kraak, M., MacEachren, A., Peuquet, D., 1997. Assessing the effectiveness of temporal legends in environmental visualization. In: *Proceedings of GIS/ILIS*, pp. 677–685.
- El-Shaarawi, A., Piegorisch, W., 2002. *Encyclopedia of Environmetrics*, 1st ed. John Wiley & Sons Inc., Sussex, England.
- Esquerdo, J., Junior, J., Antunes, J., 2009. Uso de perfis multi-tempoais de NDVI/AVHRR no acompanhamento da cultura da soja no oeste do Paraná. In: *Simpósio Brasileiro de Sensoriamento Remoto*. No. 1973. INPE, Natal, Brazil, pp. 145–150.
- Fayyad, U., Shapiro, G., Smyth, P., 1996. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM* 39 (11), 27–34.
- Fayyad, U., Stolorz, P., 1997. Data mining and KDD: promise and challenges. *Future Generation Computer Systems* 13, 99–115.
- Ferraz, S., Vettorazzi, C., Theobald, D.D., Ballester, M., 2005. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondonia, Brazil: assessment and future scenarios. *Forest Ecology and Management* 204 (January (1)), 69–85.
- Foody, G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80 (April (1)), 185–201.
- Forman, R., 1995. *Land Mosaics: The Ecology of Landscapes and Regions*, 1st ed. Cambridge University Press, Cambridge, England.
- Freitas, R., Arai, E., Adami, M., Souza, A., Sato, F., Shimabukuro, Y., Rosa, R., Anderson, L., Rudorff, B., 2011. Virtual laboratory of remote sensing series: visualization of MODIS EVI2 data set over South America. *Journal of Computational Interdisciplinary Sciences* 2, 57–64.
- Freitas, R., Shimabukuro, Y., 2008. Combining wavelets and linear spectral mixture model for MODIS satellite sensor time-series analysis. *Journal of Computational Interdisciplinary Sciences* 1 (1), 51–56.
- Frohn, R., Hao, Y., 2006. Landscape metric performance in analyzing two decades of deforestation in the Amazon Basin of Rondonia, Brazil. *Remote Sensing of Environment* 100 (January (2)), 237–251.
- Gamanya, R., Demayer, P., Dedapper, M., 2007. An automated satellite image classification design using object-oriented segmentation algorithms: a move

- towards standardization. *Expert Systems with Applications* 32 (February (2)), 616–624.
- Gavlak, A., Escada, M., Monteiro, A., 2011. Dinâmica de padrões de mudança de uso e cobertura da terra na região do Distrito Florestal Sustentável da BR-163. In: *Anais XV Simpósio Brasileiro de Sensoriamento Remoto*. INPE, Curitiba, Brazil, pp. 6152–6160.
- Goodchild, M., 2004. GIScience, geography, form, and process. In: *Association of American Geographers*, vol. 94. Blackwell Publishing, Oxford, UK, pp. 709–714.
- Groom, G., Múcher, C., Ihse, M., Wrbka, T., 2006. Remote Sensing in landscape ecology: experiences and perspectives in a European context. *Landscape Ecology* 21 (April (3)), 391–408.
- Han, J., Kamber, M., 2008. *Data Mining: Concepts and Techniques*. Technical Reports, University of Illinois at Urbana-Champaign. URL: (<http://books.google.com/books?id=Afl0t-YzOrEC>).
- Haralick, R., Shapiro, L., 1985. Image segmentation techniques. *Applications of Artificial Intelligence II* 548, 2–9.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition Springer, Stanford, CA. URL: (<http://books.google.com/books?id=tVlJmNS3Ob8C>).
- Hay, G., Castilla, G., 2008. Geographic object-based image analysis (GEOBIA): a new name for a new discipline. In: Blaschke, T., Lang, S., Hay, G. (Eds.), *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*. Springer-Verlag, Berlin, Heidelberg, pp. 75–89. (Chapter 1.4).
- Hornsby, K., Egenhofer, M., Hayes, P., 1999. Modeling cyclic change. *Advances in Conceptual Modeling* 1727 (1), 98–109.
- Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment* 83 (November (1–2)), 195–213.
- Hüttich, C., Gessner, U., Herold, M., Strohbach, B., Schmidt, M., Keil, M., Dech, S., 2009. On the suitability of MODIS time series metrics to map vegetation types in dry savanna ecosystems: a case study in the Kalahari of NE Namibia. *Remote Sensing* 1 (September (4)), 620–643. URL: (<http://www.mdpi.com/2072-4292/1/4/620/>).
- Imbernon, J., Branthomme, A., 2001. Characterization of landscape patterns of deforestation in tropical rain forests. *International Journal of Remote Sensing* 22 (June (9)), 1753–1765.
- INPE, 2012a. Deforestation estimates in the Brazilian Amazon. URL: (<http://www.obt.inpe.br/prodes/>).
- INPE, 2012b. TerraView. URL: (<http://www.dpi.inpe.br/terrawiew/>).
- ITT, 2008. *ENVI Feature Extraction Module User's Guide*. Exelis Visual Information Solutions, Gilching, Germany.
- Jiang, Z., Huete, A., Didan, K., Miura, T., 2008. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment* 112 (October (10)), 3833–3845.
- Körting, T., Dutra, L., Fonseca, L., 2011a. A resegmentation approach for detecting rectangular objects in high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters* 8 (July (4)), 621–625.
- Körting, T., Fonseca, L., Câmara, G., 2011b. A geographical approach to self-organizing maps algorithm applied to image segmentation. *Lecture Notes in Computer Science* 6915, 162–170.
- Körting, T., Fonseca, L., Escada, M., Silva, F., Silva, M., December 2008. GeoDMA—a novel system for spatial data mining. In: *2008 IEEE International Conference on Data Mining Workshops*, pp. 975–978.
- Lackner, M., Conway, T., 2008. Determining land-use information from land cover through an object-oriented classification of IKONOS imagery. *Canadian Journal of Remote Sensing* 34 (2), 77–92.
- Lang, S., 2008. Object-based image analysis for remote sensing applications: modeling reality-dealing with complexity. In: *Object-Based Image Analysis*. Springer, New York, pp. 3–27 (Chapter 1.1).
- Lang, S., Tiede, D., Developer, D., 2007. *Definiens Developer*. GIS Business 9, 34–37.
- Lewinski, S., Bochenek, Z., 2008. Rule-based classification of SPOT imagery using object-oriented approach for detailed land cover mapping. In: *Proceedings of the 28th EARSeL Symposium*. No. June. EARSeL, Istanbul, Turkey.
- McCauley, S., Goetz, S., 2004. Mapping residential density patterns using multi-temporal Landsat data and a decision-tree classifier. *International Journal of Remote Sensing* 25 (March (6)), 1077–1094.
- McGarigal, K., 2002. *Landscape pattern metrics*. In: *Encyclopedia of Environmetrics*. John Wiley & Sons, Sussex, England, pp. 1135–1142.
- McGarigal, K., Marks, B., 1994. FRAGSTATS. Spatial pattern analysis program for quantifying landscape structure. Version 2.0. Oregon State University, Corvallis (August).
- Meinel, G., Neubert, M., 2004. A comparison of segmentation programs for high resolution remote sensing data. *International Archives of Photogrammetry and Remote Sensing* 35 (Part B), 1097–1105.
- Meinel, G., Neubert, M., Reeder, J., 2001. The potential use of very high resolution satellite data for urban areas - First experiences with IKONOS data, their classification and application in urban planning and environmental monitoring. In: *Remote Sensing of Urban Areas/Fernerkundung in urbanen Räumen* (=Regensburger Geographische Schriften, Heft 35). Regensburg, pp. 196–205.
- Metzger, J., Martensen, A., Dixo, M., Bernacci, L., Ribeiro, M., Teixeira, A., Pardini, R., 2009. Time-lag in biological responses to landscape changes in a highly dynamic Atlantic forest region. *Biological Conservation* 142 (6), 1166–1177.
- Novack, T., Esch, T., Kux, H., Stilla, U., 2011. Machine learning comparison between WorldView-2 and QuickBird-2 simulated imagery regarding object-based urban land cover classification. *Remote Sensing* 3 (October (10)), 2263–2282. URL: (<http://www.mdpi.com/2072-4292/3/10/2263/>).
- Openshaw, S., 1999. Geographical data mining: key design issues. In: *Proceedings of GeoComputation* 99.
- Pacifici, F., Chini, M., Emery, W., 2009. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment* 113 (6), 1276–1292. URL: (<http://dx.doi.org/10.1016/j.rse.2009.02.014>).
- Pettorelli, N., Vik, J., Mysterud, A., Gaillard, J., Tucker, C., Stenseth, N., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology and Evolution* 20 (9), 503–510. URL: (<http://www.ncbi.nlm.nih.gov/pubmed/16701427>).
- Pinho, C., Fonseca, L., Körting, T., Almeida, C., Kux, H., 2012. Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis. *International Journal of Remote Sensing* 33 (19), 5973–5995.
- Pinho, C., Silva, F., Fonseca, L., Monteiro, A., 2008. Intra-urban land cover classification from high-resolution images using the C4.5 algorithm. In: *ISPRS Congress Beijing*, vol. 7.
- Quinlan, J., 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA. URL: (<http://books.google.com/books?id=HExncpbjYroC>).
- Ribeiro, M., Metzger, J., Ponzoni, F., Martensen, A., Hirota, M., 2009. The Brazilian Atlantic forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation* 142 (6), 1141–1153.
- Rubinstein, R., Kroese, D., 2008. *Simulation and the Monte Carlo Method*. Wiley-Interscience, NJ.
- Saito, E., Fonseca, L., Escada, M., Körting, T., 2011. Effects of changes in scale of deforestation patterns in Amazon. *Brazilian Journal of Cartography* 4.
- Saito, E., Körting, T., Fonseca, L., Escada, M., 2010. Mineração em dados espaciais de desmatamento do prodes utilizando métricas da paisagem caso de estudo município de Novo Progresso- PA. In: *III Simposio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*. Recife, Brazil.
- Silva, M., Câmara, G., Escada, M., Souza, R., 2008. Remote-sensing image mining: detecting agents of land-use change in tropical forest areas. *International Journal of Remote Sensing* 29, 4803–4822.
- Silva, M., Câmara, G., Souza, R., Valeriano, D., Escada, M., 2005. Mining patterns of change in remote sensing image databases. In: *The Fifth IEEE International Conference on Data Mining*. Citeseer, New Orleans, Louisiana, USA.
- Small, C., 2011. Spatiotemporal dimensionality and time-space characterization of vegetation phenology from multitemporal MODIS EVI. In: *2011 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images (Multi-Temp)*. IEEE, pp. 65–68.
- Smith, B., 1995. On drawing lines on a map. In: Frank, A., Kuhn, W., Mark, D. (Eds.), *Proceedings of COSIT (Conference on Spatial Information Theory)*. Springer Verlag, Semmering, Austria, pp. 475–484.
- Southworth, J., Nagendra, H., Tucker, C., 2002. Fragmentation of a landscape: incorporating landscape metrics into satellite analyses of land-cover change. *Landscape* 27 (3), 253–269.
- Stein, A., Hamm, N., Ye, Q., 2009. Handling uncertainties in image mining for remote sensing studies. *International Journal of Remote Sensing* 30 (20), 5365–5382.
- Steiniger, S., Hay, G., 2009. Free and open source geographic information tools for landscape ecology. *Ecological Informatics* 4 (September (4)), 183–195.
- Stojmenovic, M., Nayak, A., Zunic, J., 2008. Measuring linearity of planar point sets. *Pattern Recognition* 41 (August (8)), 2503–2511.
- Tan, P., Steinbach, M., Kumar, V., Potter, C., Klooster, S., Torregrosa, A., 2001a. Clustering earth science data: goals, issues and results. In: *Proceedings of the Fourth KDD Workshop on Mining Scientific Datasets*.
- Tan, P., Steinbach, M., Kumar, V., Potter, C., Klooster, S., Torregrosa, A., 2001b. Finding spatio-temporal patterns in earth science data. *Earth Science*, 1–12.
- Turner, M., 2005. *Landscape ecology: what is the state of the science?*. *Annual Review of Ecology, Evolution, and Systematics* 36 (December (1)), 319–344.
- Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment* 114 (January (1)), 106–115.
- Vermote, E., Saleous, N., Justice, C., 2002. Atmospheric correction of MODIS data in the visible to middle infrared: first results. *Remote Sensing of Environment* 83, 97–111.
- Wassenberg, J., Middelmann, W., Sanders, P., 2009. An efficient parallel algorithm for graph-based image segmentation. In: *Computer Analysis of Images and Patterns*. Springer, Muenster, Germany, pp. 1003–1010.
- Wiens, J., 1976. Population responses to patchy environments. *Annual Review of Ecology and Systematics* 7, 81–120.
- Woodcock, C., Strahler, A., 1987. The factor of scale in remote sensing. *Remote Sensing of Environment* 21 (April (3)), 311–332.