

Remote-sensing image mining: detecting agents of land-use change in tropical forest areas

MARCELINO PEREIRA DOS SANTOS SILVA^{†‡}, GILBERTO CÂMARA^{*‡},
MARIA ISABEL SOBRAL ESCADA[†] and RICARDO CARTAXO MODESTO
DE SOUZA[‡]

[†]UERN—Rio Grande do Norte State University, BR 110, km 46, 59625-620, Mossoró,
RN, Brazil

[‡]INPE—National Institute for Space Research, PO Box 515, 12201-097, São José dos
Campos, SP, Brazil

(Received 30 December 2005; in final form 4 July 2007)

Land remote-sensing images are the primary means of assessing land change. There have been major land changes in the planet in the last decades, especially in tropical forest areas. Identifying the agents of deforestation is important for establishing public policies that can help preserve the environment. This paper proposes a method for detecting the agents of land change in remote-sensing image databases. We associate each land-change pattern, detected in a remote-sensing image, to one of the agents of change. The proposed method uses a decision-tree classifier to describe shapes found in land-use maps extracted from remote-sensing images and then associates these shape descriptions to the different types of social agents involved in land-use change. We support our proposal with two case studies for detecting land-change agents in Amazonia, using the remote-sensing image database of the Brazilian National Institute for Space Research (INPE).

1. Introduction

Land remote-sensing images are the primary means of assessment of land change worldwide. From these images, we know that the planet has experienced major land changes in the last decades, especially in tropical forest areas (Lambin *et al.* 2003). Since it is the world's largest tropical forest, deforestation in the Amazonia rainforest is important for global land change and a significant contributor to the global carbon budget as well as having known impacts on climate (Shukla *et al.* 1990) and biodiversity (Fearnside 2002). The Brazilian National Institute for Space Research (INPE) uses satellite images to provide yearly assessments of the deforestation in Amazonia. According to INPE's estimates, close to 200 000 km² of forest was cut in Amazonia in the period from 1995 to 2005 (INPE 2005). Given the extent of this deforestation, it is important to determine the agents of deforestation for setting up public policies that can help preserve the environment.

There is a consensus that land change has multiple causes and that local agents play a major role in controlling the extent of change (Lambin *et al.* 2003). Land-change agents in Amazonia include rubber-tappers, cattle ranchers, large

*Corresponding author. Email: gilberto.camara@inpe.br

agricultural farmers, small-scale landowners, and government-induced settlements (Alves *et al.* 2003). There is a dynamic interplay between these agents. Cattle ranchers may expand into areas formerly occupied by rubber-tappers. Government programmes may confiscate land from farmers and give it out to settlers. Thus, both market forces and public policies influence land change in Amazonia (Andersen and Reis 1997, Pfaff 1999, Perz and Skole 2003).

To gain a better understanding of land change in Amazonia, we need to assess the role and the spatial organization of the different agents involved in land change. We need to associate each land-change patch, detected in a remote-sensing image, to one of the agents of change. Extensive fieldwork indicates that the different agents involved in land-use change (small-scale farmers, large plantations, cattle ranchers) can be distinguished by their different spatial patterns of land use (Lambin *et al.* 2003). These patterns evolve in time; new small settlements emerge, and large farms increase their agricultural area at the expense of the forest. In these and similar cases, patterns of land-use change will have similar spectral signatures (Mas 1999). Therefore, we need techniques that are able to distinguish patterns of land-use change based on their shapes and spatial arrangements.

Given this motivation, this paper proposes a method for detecting agents of change in land remote-sensing image databases. The method starts by identifying the different types of land-change agents in a selected area. Next, it builds a training set of land-change patches, where each patch is associated with one type of local agent. Then, it uses landscape ecology metrics to label the patches and a decision tree to classify them. Our approach builds on earlier works by our research group (Câmara *et al.* 2001), and this paper is an extended and revised version of earlier results (Silva *et al.* 2005). Since the method uses landscape ecology metrics, we discuss previous work on metrics for land-change modelling in §2. We describe our method for mining land-use agents in remote-sensing image databases in §3. §4 presents two case studies for detecting land-use agents in Amazonia from INPE's remote-sensing image database.

2. Background

The Amazon deforestation surveys carried out by INPE use change-detection techniques based on spectral mixture models of remote-sensing images (Shimabukuro *et al.* 1998). INPE has built a large spatial database of yearly land changes in the region since 1997. This database has proven to be useful for researchers that study the causes and factors of Amazon deforestation (Alves 2002, Alves *et al.* 2003, Escada *et al.* 2005b, Ewers and Laurence 2006). In the present work, we use patch metrics from landscape ecology to classify the land-change patches of INPE's deforestation database.

To put our work in context, in this section we discuss previous works that use landscape-ecology metrics for land-change modelling, with an emphasis on Amazonia. Mertens and Lambin (1997) identify three types of deforestation patterns in Cameroon. The authors link landscape metrics to the frequency of deforestation. Peralta and Mather (2000) use multitemporal Landsat satellite imagery to analyse forest metrics (lacunarity, patchiness, and area-perimeter), where each metric is associated with a specific type of land use. Imbernon and Branthomme (2001) model the landscape dynamics on a Brazilian site, by measuring the percentage of forest cover and the resulting fragmentation. Southworth *et al.* (2002) used metrics of land-cover change to infer patterns of land-use change in

Honduras with Landsat TM imagery from 1987, 1991, and 1996. Oliveira Filho and Metzger (2006) tested if abrupt changes (that is, thresholds) could be detected by landscape structure indices in real and simulated landscapes. They compared three different deforestation patterns in the Brazilian Amazon: small properties regularly spread along roads (fishbone), irregularly scattered small properties (independent settlements), and large properties. The focus of their work was in habitat loss and conservation biology. Ewers and Laurence (2006) used fractal dimension as a landscape index to examine the area–perimeter ratio of deforested areas in the Brazilian Amazon. They tested scale invariance in deforestation patterns comparing different maps obtained from AVHRR and Landsat images. The spatial pattern of deforestation differed between small and large clearings. The authors found that different deforestation processes leave distinctive footprint scales.

As an alternative to landscape ecology methods, previous papers (Kuhnert *et al.* 2006) propose that land-use patterns may be understood as a result of self-organization principles. In particular, researchers have identified that land-use patterns in the Amazonian region of Ecuador follow a power-law distribution (Malanson *et al.* 2006). One of the important concerns in power-law distribution is how to estimate its coefficients. One useful technique for such an estimation is the detrended fluctuation analysis (DFA), which has already proved its usefulness in several complex problems like the total ozone content (Varotsos 2005a, b).

The review indicates that many previous works have used landscape ecology metrics to analyse land-change patches extracted from classified remote-sensing images. These metrics were used either to describe landscape structure or to provide as an overall assessment of factors associated with deforestation. Past authors have not proposed a systematic method for associating the pattern shapes with the local agents. We have not come across works where the landscape ecology metrics are used by an automated classifier that associates land-change patches with local agents, as we propose in this paper.

3. Methods

3.1 General perspective

To gain a better understanding of land-change using a remote-sensing image database, researchers would like to explore the database with questions such as: What are the different land-use agents present in the database? When did a certain land-use agent emerge? What are the dominant land-use agents for each region? How do agents emerge and change in time? To answer these and similar questions, we propose a method that associates land-change patches with causative agents of change.

The starting point of the method is a sequence of medium-resolution (LANDSAT-class) images. These images are then classified by a segmentation procedure, which detects the changes that have occurred between two consecutive images. The result is a set of land-change maps. The earliest map of the sequence contains a set of polygons covering the whole study area. Each polygon is labelled to a land-cover class. Each of the other maps contains only the areas that have changed between two consecutive images. For example, INPE's Amazonia deforestation database starts at 1997 and has yearly updates. The 1997 map contains polygons of one of four classes: 'forest', 'non-forest', 'deforestation', and 'rivers'. The other maps contain only the regions of change since the previous map. We refer to these polygons as 'land-change patches'. These land-change patches are the focus of attention of our data-mining technique (figure 1).

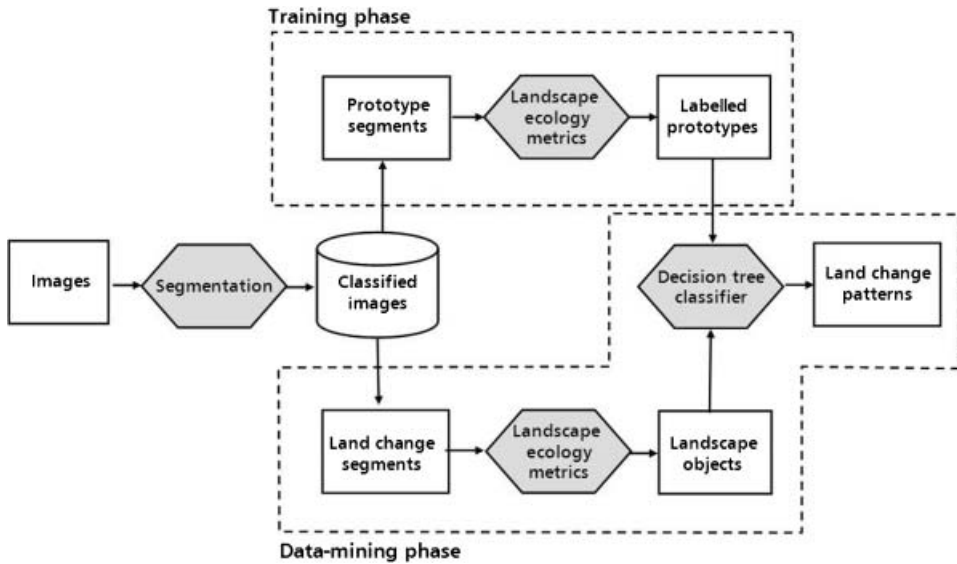


Figure 1. Proposed method for remote sensing image mining.

The proposed data-mining technique uses the idea of a *land-change patch*. A *land-change patch* is a closed region detected in a remote-sensing image and associated with a change in land cover. The first part (shown in the upper part) is the training procedure. Based on established knowledge about the agents that cause land change, the analyst defines a *spatial pattern typology* according to an application domain. The analyst selects a training set of land-change patches and assigns a label to each one, choosing among the different types of spatial patterns. Each patch then has a set of attributes, using landscape-ecology metrics. The result is a training set of patches, where each member has a label and a set of numerical characteristics. The training set is used to build a decision-tree classifier, which distinguishes each type of land-change patch based on its patch metrics.

The second part of the method (shown in the lower part of figure 1) is the data-mining procedure. The analyst calculates a set of numerical attributes of all land-change patches using landscape-ecology metrics. The decision-tree classifier (built in the training procedure) maps each land-change patch to one of the user-defined spatial patterns. Finally, the user analyses the spatio-temporal trends of the resulting land-change patterns. For example, the results may indicate an increase in cattle ranching during the last 5 years in a specific area. In what follows, we present each of these steps in detail.

3.2 Defining a spatial pattern typology

The first phase of the method calls for defining a spatial pattern typology for the study area. Each of these patterns matches a specific real-world action. Once the user fixes a typology of spatial patterns, they select prototypical examples in the images. These prototypes will be the training set of the patch classifier. For example, Mertens and Lambin (1997) propose a typology of the land-use pattern associated with deforestation in tropical forests, extracted from remote-sensing images (figure 2). Their typology includes *corridor* (commonly associated with riverside and roadside colonization), *diffuse* (related to smallholder subsistence agriculture), *fishbone* (typical

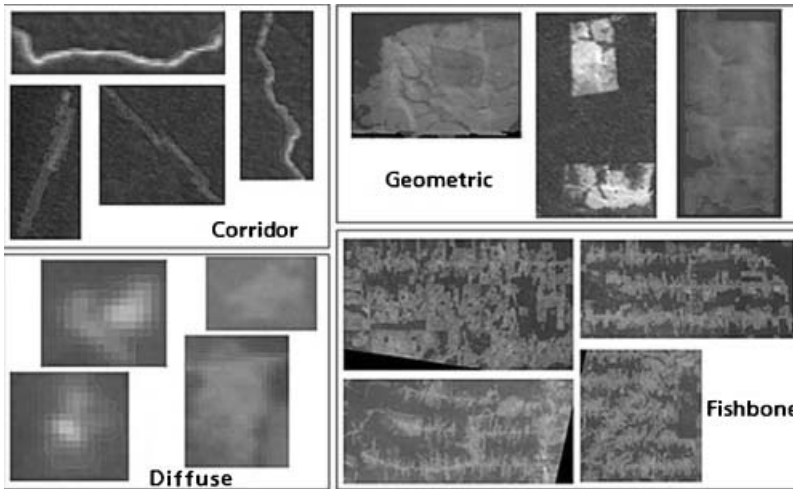


Figure 2. Examples of patterns of tropical deforestation proposed by Mertens and Lambin (1997) in the Brazilian Amazonia: corridor, diffuse, fishbone, and geometric.

of planned settlement schemes), and *geometric* (linked to large-scale clearings). Figure 2 shows an example of this typology for a region in Amazonia. There are four prototypes for each pattern, extracted from a LANDSAT TM image.

The spatial patterns typology proposed by Mertens and Lambin (1997) aim to determine the agents involved in tropical deforestation in Cameroon. This typology may not be satisfactory to describe deforestation in other regions. Associating land-change patterns with social agents in a specific region needs an understanding of occupation history, economic, social, and environmental constraints. In the Brazilian Amazonia, there are many fishbone patterns associated with planned settlements, similar to the patterns proposed by Mertens and Lambin (1997). However, there are other spatial patterns associated with settlements, since the Brazilian government used different spatial arrangements to organize colonist land parcels in planned settlement schemes. Colonist land parcels use different spatial arrangements, including fishbone, radial, corridors, and dendritic patterns (Batistella *et al.* 2003, Escada *et al.* 2005a).

3.3 Image segmentation

To extract patterns from remote-sensing images, we use a segmentation algorithm to partition the image into regions that are spatially continuous, disjoint, and homogenous. In our examples, we have adopted the region-growing algorithm developed by INPE (Bins *et al.* 1996) and included in the SPRING software (Câmara *et al.* 1996), which is change patches in tropical forests (Shimabukuro *et al.* 1998). INPE uses it to produce its estimates of deforestation in Amazonia (INPE, 2005). A recent survey of segmentation algorithms gave the method a favourable assessment (Meinel and Neubert 2004). However, the proposed method is not limited by this particular code. Our method can use any suitable region-growing algorithm that guarantees creating closed regions for remote-sensing images.

The SPRING algorithm uses two parameters: a *similarity threshold* and an *area threshold*. It starts by comparing neighbouring pixels and merging them into regions if they are similar. The algorithm then tries iteratively to merge the resulting regions.

Two neighbouring regions, R_i and R_j , are merged if they satisfy the following conditions:

1. threshold condition: $d(R_i, R_j) \leq T$;
2. neighbourhood condition 1: $R_j \in N(R_i)$ and $d(R_i, R_j) \leq d(R_k, R_i)$, $R_k \in N(R_i)$;
3. neighbourhood condition 2: $R_i \in N(R_j)$ and $d(R_i, R_j) \leq d(R_k, R_j)$, $R_k \in N(R_j)$.

In these conditions, T is the *similarity threshold*, $d(R_i, R_j)$ is the Euclidean distance between the mean grey levels of the regions, and $N(R)$ is the set of neighbouring regions of region R . Regions smaller than the chosen *area threshold* are merged with its most similar neighbour. The results of the segmentation are sensitive to the choice of similarity and area thresholds. Low values of area threshold result in excessive partitioning, producing a confusing visual picture of the regions. High values of similarity threshold force the union of spectrally distinct regions, resulting in undersegmentation. In addition, the right thresholds vary depending on the spectral range of the image.

3.4 Definition of patch metrics for land-change patches

We use patch metrics from landscape ecology (Turner 1989) to select the attributes that distinguish the different types of land-change patches. Landscape-ecology theory defines a landscape as an area of land containing a mosaic of patches. It considers that land patterns strongly influence ecological processes and proposes metrics for the geometrical and spatial properties of patches (McGarigal 2002). Patch metrics refers to the spatial character and arrangement, position, or orientation of patches within the landscape. Our method uses the patch metrics proposed by the FRAGSTATS software (Spatial Pattern Analysis Program for Categorical Maps) (McGarigal and Marks 1995), which include the following metrics, where p_{ij} is the perimeter (m), and a_{ij} is the area (m^2) of region ij :

- Perimeter (m):

$$\text{PERIM} = p_{ij}. \quad (1)$$

- Area (ha):

$$\text{AREA} = (a_{ij}/10\,000). \quad (2)$$

- PARA, perimeter–area ratio, a measure of shape complexity:

$$\text{PARA} = \frac{p_{ij}}{a_{ij}}. \quad (3)$$

- Shape, shape compactness index, calculated by the patch perimeter p_{ij} divided by $p_{ij \text{ min}}$, which is the minimum perimeter possible for a maximally compact patch of the matching patch area. It is equal to 1 when the region is a square and grows according to the region's irregularity.

$$\text{SHAPE} = \frac{p_{ij}}{p_{ij \text{ min}}}. \quad (4)$$

- Frac, fractal dimension index for measuring shape complexity, which approaches 1 for shapes with simple perimeters such as squares, and approaches 2 for complex shapes:

$$FRAC = \frac{2 \ln(0.25p_{ij})}{\ln a_{ij}}. \quad (5)$$

- Circle (circumscribing circle), equal to 0 for circular patches and closer to 1 for elongated ones, which is calculated based on the patch area a_{ij} and is the area a_{ij}^s of the smallest circumscribing circle for region ij :

$$CIRCLE = 1 - \left[\frac{a_{ij}}{a_{ij}^s} \right]. \quad (6)$$

- Contig assesses the spatial contiguity of a patch. It is quantified by convolving an image in which the pixels of the patch are assigned a value of 1 and all other pixels are set to 0 with a 3×3 mask with the following values:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

This combination of integer values weights orthogonally contiguous pixels more heavily than diagonally contiguous pixels. For each pixel r in the patch ij , the value c_{ijr} is calculated by multiplying the 3×3 mask by a 3×3 window of the binary image centred on the pixel. All values are added and divided by n , the total number of pixels in the patch. The result is then normalized by the sum v of the values of the mask (equal to 13). A value of 1 is subtracted from both the numerator and denominator for the index to vary from 0 to 1. The CONTIG index equals 0 for a one-pixel patch and increases towards 1 as patch contiguity increases:

$$CONTIG = \frac{\sum_{r=1}^n c_{ijr}}{n} - 1. \quad (7)$$

- Gyrate (radius of gyration) is a measure of patch extent, influenced by both patch size and patch compaction. For each pixel r in the patch ij , it computes the distance d_{ijr} from the pixel to the centroid (average location) of patch. This distance is then divided by the total number of pixels n in the patch:

$$GYRATE = \frac{\sum_{r=1}^n d_{ijr}}{n}. \quad (8)$$

The metrics for the prototype patches trains the decision-tree classifier, which selects the best combination of metrics that produces an efficient classification.

3.5 Building a decision-tree classifier

The role of the decision-tree classifier is to label each land-change patch to one of the spatial patterns associated with the land-change agents in the study area. Consider, for example, that we are using the spatial pattern typology proposed by Mertens and Lambin (1997). Then, each land-change patch will be either one of a *corridor*, *diffuse*, *fishbone*, or *geometric* pattern (figure 2).

The classifier maps the land-change patches to the spatial patterns typology. Its training set is a set of patches where each patch has a descriptive label and a set of numerical attributes. This problem can be expressed as a classification based on a decision tree that predicts correctly the value of a categorical attribute, based on numerical attributes (Witten and Frank 1999). The categorical value is the patch type, and the numerical values are the patch metrics. Our method can use any such algorithm. For the case studies, we used the C4.5 decision-tree classifier (Quinlan 1993), which builds a decision tree where:

- Each node matches a numerical attribute, and each arc is one of its possible values. A leaf of the tree specifies the expected value of the categorical attribute for the records described by the path from the root to that leaf.
- Each node is associated the numerical attribute which is *most informative* among the attributes not yet considered in the path from the root. Node information is obtained from an *entropy* measure.

The landscape-ecology metrics of the training set of land-change patches (as in figure 2) are fed into the C4.5 classification algorithm. The algorithm builds a decision tree that uses these metrics to distinguish the different types of patches. After the classifier has been properly trained, it labels the land-change patches obtained from remote-sensing images (figure 1). Figure 3 shows the decision tree for the prototype patches from figure 2.

The decision tree in figure 3 uses two metrics. The first metric is AREA (area in km²) which distinguishes the smaller types of patches (diffuse and corridor) from the larger patches (geometric and fishbone). All patches greater than 386 km² belong to one of the two latter types. The remaining distinction uses the SHAPE metric, a compactness index which is equal to 1 for a square patch. It distinguishes the diffuse

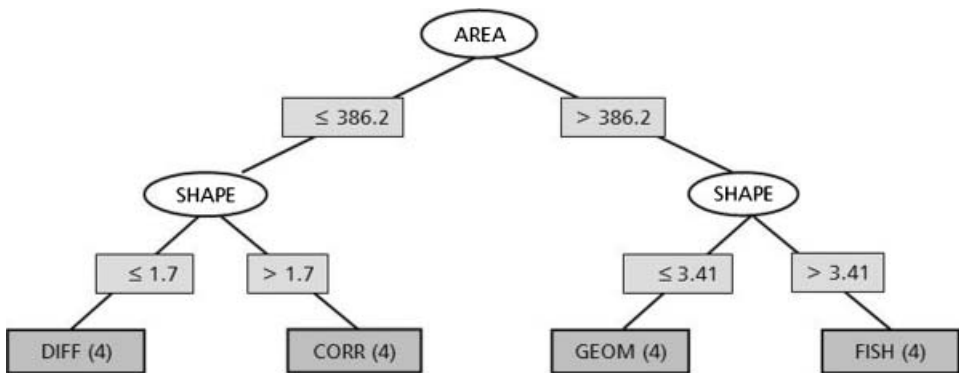


Figure 3. Decision tree for patterns in figure 3 (GEOM: geometric; FISH: fishbone; DIFF: diffuse; CORR: corridor). Metrics: area in km² (AREA) and shape compactness index (SHAPE).

patches (more compact) from the corridor patches (less compact). It also separates the geometric patches (more compact) from the fishbone patches (less compact). This decision tree was validated using the different set of test patterns. The classifier had an 81% correctness rate with the test set.

3.6 Mining the database using segmentation and a decision-tree classifier

In this section, we describe the image data-mining procedure shown in the lower part of figure 1. The decision-tree classifier associates each land-change patch with one of the pattern types defined by the user. The data-mining procedure has five steps:

1. Selecting a sequence of images from the database.
2. Segmenting and classifying the sequence of images to find land-change patches.
3. Labelling the land-change patches using the metrics described in §3.4.
4. Assigning each land-change patch to one of the user-defined patterns.
5. Analysing the temporal evolution of the patterns to evaluate the impact and influence of each land-change agent on the study area.

By identifying the spatial patterns on a time series of images, the user will be able to evaluate the emergence and evolution of different types of change. Each spatial pattern is associated with a specific land-change agent. Therefore, the comparison between spatial patterns of images at the same location in different times allows new insights into the agents that bring about change.

4. Case studies: results and discussion

We used the method described in the previous section to gain a better understanding of land change in Amazonia. We selected two case studies. The first study case is the *Xingu-Iriri* region in the *Pará* state. There, deforestation has increased in the last 5 years, associated with unplanned occupation and cattle ranching. The second study case is a planned rural settlement in the *Vale do Anari* municipality in the state of Rondônia. In both cases, our work used LANDSAT imagery at 30-m spatial resolution, acquired from INPE's archives.

4.1 Xingu-Iriri case study

The Xingu-Iriri region is a large area in the state of Pará, where squatters seized a considerable area of public land by illegal procedures in recent years (Becker 2004, Escada *et al.* 2005b). The area covers around 150 000 km², or 10% of the state of Pará. This region includes parts of *São Félix do Xingu* and *Altamira* municipalities. It stands between two important rivers of Pará: the Xingu river, one of the largest tributaries of the Amazonas river, and the Iriri river. São Félix do Xingu is a region with many violent deaths due to land conflicts. This city has one of the largest annual rates of deforestation in Amazonia and has 10% of the cattle of Pará state. Cattle raising in São Félix grew by 780% from 1997 to 2004. Deforestation in the region has two important agents: migrants that have settled in small areas, and large cattle ranchers, many of whom have occupied land illegally. Although partly illegal, cattle ranching is an organized business, and farmers have access to regional and national markets. Figure 4 shows the study area.

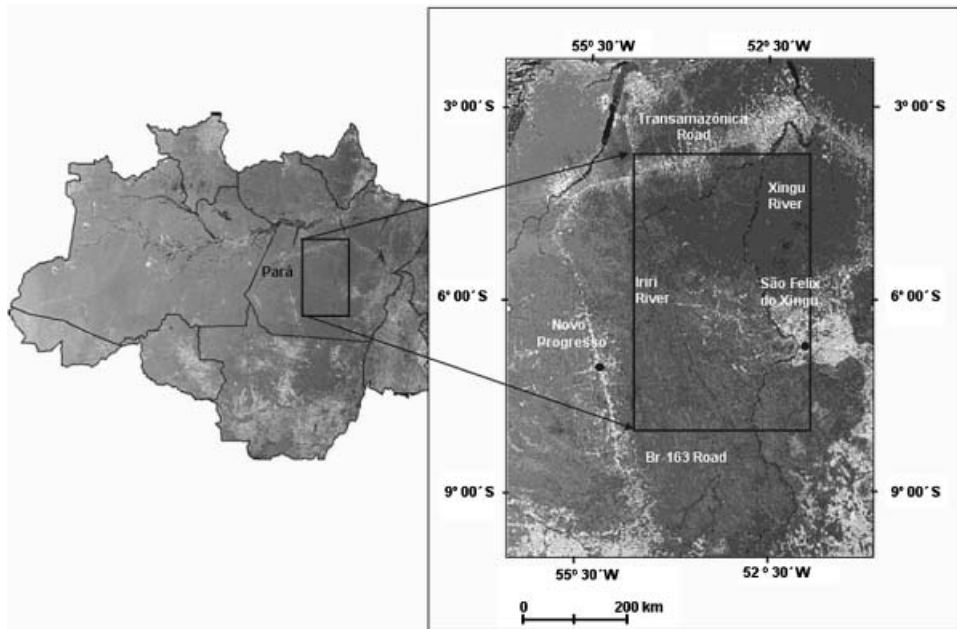


Figure 4. Location of the study area. The Brazilian Amazonia is on the left, and the Xingu-Iriri watershed in the state of Pará is on the right. Light-coloured areas indicate deforestation.

We wanted to assess the behaviour of different types of farmers from 1997 to 2004 and the spatial organization of the farms. To this end, we identified five types of agents in the Xingu-Iriri region:

- Small households associated with migrant families, who live on subsistence agriculture or work for the farmers. Their land-use pattern is associated with earlier roadside colonization and shows up as linear patterns in the classified maps. These households appear as *small linear patches*.
- Very small farmers and family households that live out near the main roads or close to population settlements. We associate these ranchers with irregular land-use patterns of size less than 35 ha. These households appear as *very small irregular patches*.
- Small cattle ranchers that live near to roads or to settlements. We associate these ranchers with irregular land-use patterns of size between 35 and 190 ha. Their land-use patterns show up in the maps as *small irregular near-road patches*.
- Farmers with medium-sized land-use patterns (190–900 ha) that are isolated or close to secondary roads. We associate these households with *medium irregular near-road patches*.
- Farmers with large-sized properties (more than 900 ha) that are usually isolated and located close to rivers and secondary roads, and that show up in the maps as *large geometric patches*.

Table 1 presents the typology for land-use agents, and figure 5 shows examples of the five spatial patterns (linear, very small irregular, small irregular, medium irregular, and large geometric).

Table 1. Typology of land-change actors in the Xingu-Iriri region.

Land-use patterns	Spatial distribution	Clearing size	Actors	Main land use	Description
Linear (LIN)	Roadside	Variable	Small households	Subsistence agriculture	Roadside clearings, following main roads
Small isolated irregular (SMA)	Near main settlements and main roads	Very small (<35 ha)	Small farmers	Family labour and cattle ranching	Near main roads and settlements up to 10 km
Irregular near-road small (IRR)	Near main settlements and main roads	Small (35–190 ha)	Small farmers	Cattle ranching	Associated with small family households
Irregular near-road medium (MED)	Isolated or near secondary roads	190–900 ha	Medium farmers	Cattle ranching	Associated with medium to large farms
Large geometric (LAR)	Isolated or at the end of secondary roads	Large (>900 ha)	Large farmers	Cattle ranching	Isolated; may have airstrips

We produced deforestation maps by segmentation and classification of LANDSAT imagery (INPE 2005) for the period 1997–2004, with 30-m spatial resolution. We extracted prototypical land-change patches from the deforestation maps (some examples appear in figure 5). We used one image for the 3-year period 1997–2000, and one image a year from 2000 to 2004. We used 85 instances of land-change patches to train the decision-tree classifier. The C4.5 algorithm (Quinlan 1993) built a decision tree with four metrics and four levels (figure 6).

The decision tree in figure 6 uses the radius of gyration metric (GYRATE) to distinguish the very small patches from the rest. *Very small* patches have a small extent, which results in a smaller radius of gyration. Then, it uses the *AREA* metric (area in km²) which separates the other small patches (all *small irregular* patches and some *linear* patches) from the larger ones. The *SHAPE* metric (compactness index) then distinguishes *small irregular* patches from the *linear* ones. To distinguish the larger patches, the tree uses two measures. First, the radius of gyration metric (GYRATE) is used again to identify the medium-sized patches. The *large geometric* patches and the large *linear* patches are split using the *PARA* (perimeter-to-area ratio). Geometric patches have a lower ratio than linear patches. We tested the classifier's behaviour using leave-one-out cross-validation. This method takes a single observation from the original sample as the validation data, and the remaining observations as the training data. Each observation in the sample is used once as the validation data. The cross-validation showed a 94% accuracy rate.



Figure 5. Spatial patterns in the Xingu-Iriri region: (from left to right) linear, very small irregular, small irregular, medium irregular, large geometric.

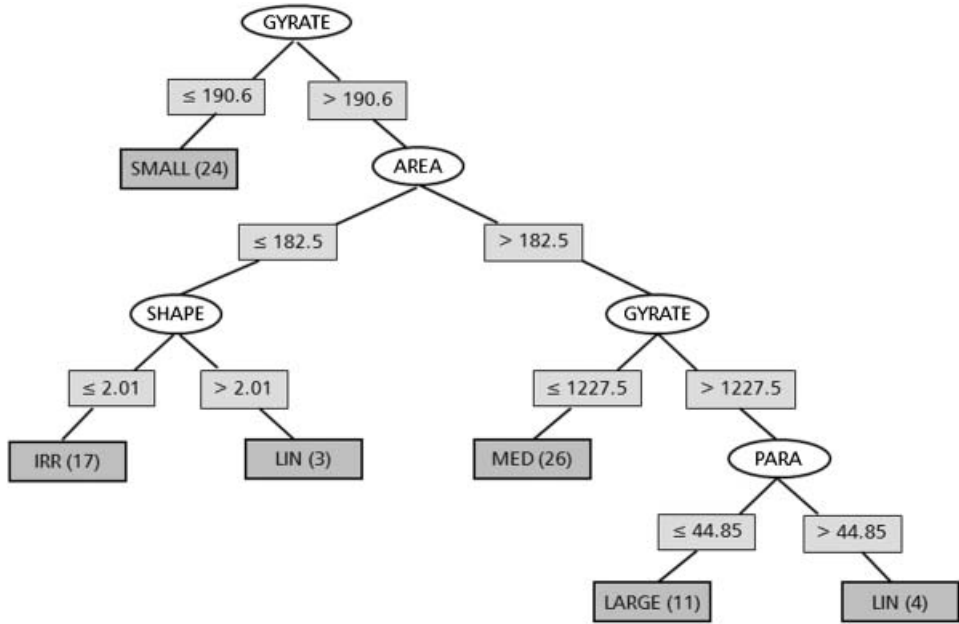


Figure 6. Decision tree for Xingu-Iriri spatial patterns. The metrics are: radius of gyration (GYRATE), perimeter/area ratio (PARA), area (AREA), and shape compactness index (SHAPE).

Using the decision tree, we built a distribution of types of clearing patterns on the Xingu-Iriri region from 1997 to 2004 (figure 7). It shows how human occupation evolved in this region. The deforestation rate started to increase after 2001 and reached a peak of 40,000 ha in the period 2001–2002. In 1997, linear patterns dominated, associated with road construction and roadside farm clearings. The most important contribution to deforestation rates from 2001 to 2004 came from large and medium geometric patterns, associated with large and medium farms. The deforestation patterns show a trend towards land concentration, where large farms

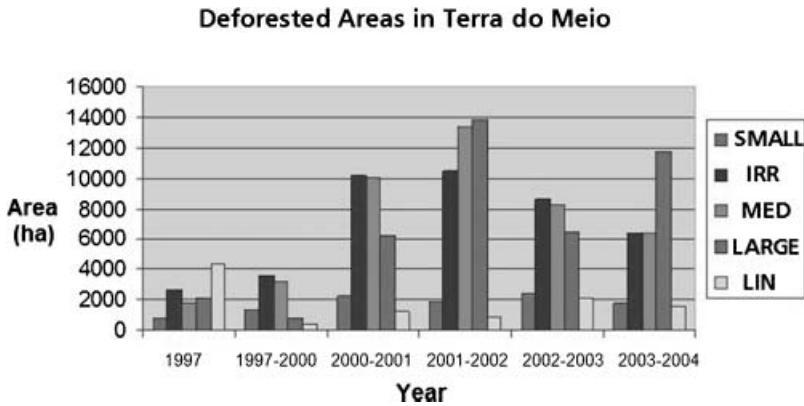


Figure 7. Distribution of deforestation in the Xingu-Iriri region (1997–2004) by patch types: linear (LIN), very small irregular (SMALL), small irregular (IRR), medium irregular (MED), and large geometric (LARGE).

dominate over small settlements. We present the cumulative clearing patterns for a part of Xingu-Iriri region in figure 8. It shows that the resulting spatial arrangement has small farms and family households concentrated along main roads, and large and medium farms arranged near secondary roads and in remote places.

To validate these results, we carried out two field trips in the region in 2004 (Escada *et al.* 2005b) and in 2006 (Amaral *et al.* 2007). Due to the large size of the region, it is unfeasible to do a detailed ground survey. The Xingu-Iriri region is similar in size to Uruguay, and is greater than Austria and Switzerland put together. Also, researchers and surveyors are undesirable visitors in the area, and they cannot count on effective police protection. Therefore, our fieldwork focused on the area close to the main road that connects São Felix do Xingú to the Iriri river, called Canopus road, shown in figure 8. We interviewed local settlers and authorities, whenever possible.

The field trips allowed us to determine how the region was occupied. The Canopus Mining Company opened the so-called 'Canopus road' in the beginning of the 1980s to support cassiterite mining. Migrant families and mahogany loggers then

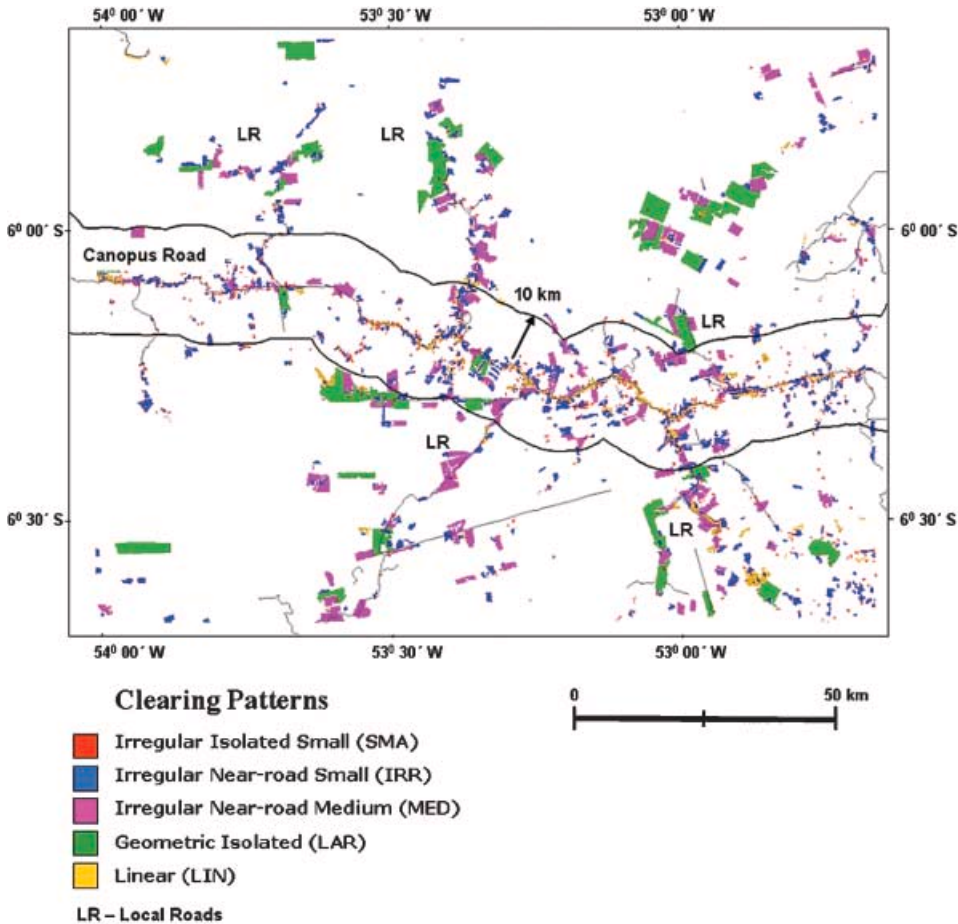


Figure 8. Cumulative clearing patterns for the Xingu-Iriri region (1997–2004) by patch type: linear (LIN), very small irregular (SMA), small irregular (IRR), medium irregular (MED), and large geometric isolated (LAR).

used the road to invade the region. The Land Institute of Pará State encouraged the occupation by giving out land parcels of 100 ha to colonists, up to 10 km from the Canopus road. In the early 1990s, some villages started to emerge along this road. Mahogany logging lasted until the end of 1990s, when all supplies had been exploited. Then, farmers and cattle ranchers entered the area using the dense road network opened by loggers (Amaral *et al.* 2007). These farmers set up large and medium farms near vicinal roads or in isolated areas, close to rivers. Large farms are located far from the main road and often have small airports.

We compared our field data with the results from the data mining. Data-mining results indicate that 61%, 68%, and 49% of the linear, small irregular, and irregular patches, respectively, lie within 10 km of the Canopus road. Also, 70% of medium irregular patches and 93% of large geometric patches are farther than 10 km from the road. These results are consistent with our fieldwork. Thus, land-occupation patterns detected by data mining are a reliable guide for describing the history of occupation in the Xingu-Iriri region.

4.2 Vale do Anari case study

The second case study used a government planned rural settlement in Vale do Anari municipality in the state of Rondônia (shown in figure 9). This settlement started in 1982, with land parcels sized around 50 ha. The study area comprises about 4000 km². We wanted to examine land concentration in Vale do Anari. Land concentration results from merging of many land parcels in a rural settlement, where one farmer buys the parcels from the original settlers. This results in farms with a medium to large size. Many studies in the literature indicate that land concentration

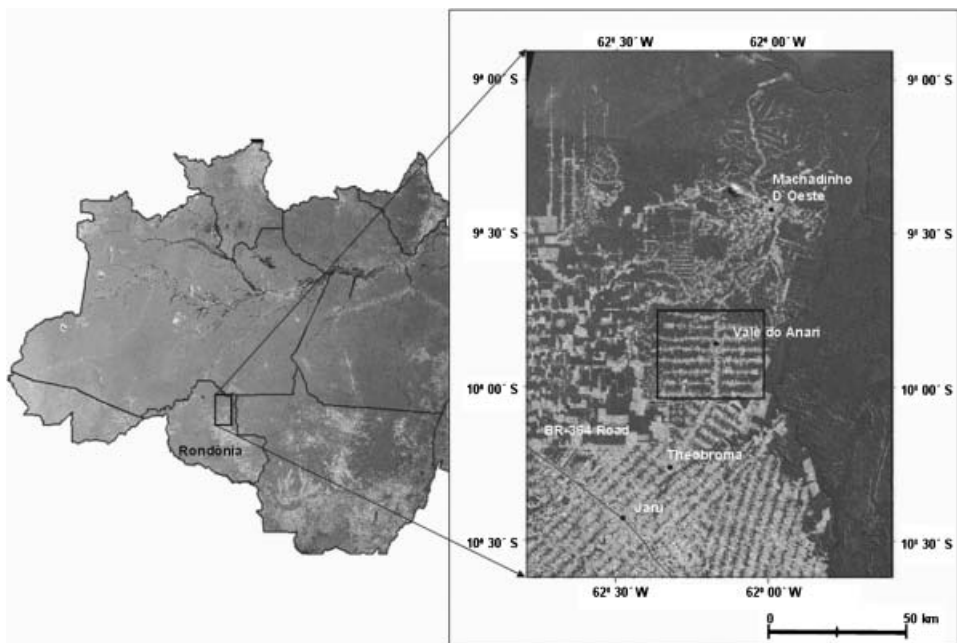


Figure 9. Location of the study area. The Brazilian Amazonia is on the left, and the Vale do Anari area in the state of Rondônia is on the right. Light-coloured areas indicate deforestation.

Table 2. Typology of land-change agents in the Vale Do Anari region.

Land-use patterns	Spatial distribution	Clearing size	Actors	Main land use	Description
Linear (LIN)	Roadside	Variable	Small households	Subsistence agriculture	Settlement parcels less than 50 ha; deforestation uses linear patterns following government planning.
Irregular (IRR)	Near main settlements and main roads	Small (<50 ha)	Small farmers	Cattle ranching and subsistence agriculture	Settlement parcels less than 50 ha; irregular clearings near roads following settlement parcels
Regular (REG)	Near main settlements and main roads	Medium and large (>50 ha)	Midsized and large farms	Cattle ranching	Patterns produced by land concentration

occurs in government settlement areas in Amazônia (Schimink and Wood 1992, Dale *et al.* 1994, Almeida and Campari 1996, Escada *et al.* 2005a). Although selling parcels in government settlements is mostly illegal, it is an established business. Settlers who are successful in using the land for farming or cattle raising buy land from those who migrate to other areas, a practice known as ‘turnover’ (Campari 2005).

We considered two types of social agents in the Vale do Anari area, associated with three spatial pattern types (summarized in table 2 and figure 10):

- Small settlement household colonists living on subsistence agriculture or small cattle ranching. Their spatial patterns show up as *linear* patterns following planned roads built during earlier stages of colonization.
- Small household colonists associated with settlement schemes living on subsistence agriculture or small cattle ranching. Their spatial patterns show up as *irregular* clearings near roads, following parcels defined by the planned settlement.
- Medium to large farmers, associated with cattle ranches larger than 50 ha. Their spatial patterns are *regular* ones, close to roads and population nuclei.

In this study, we wanted to understand how land concentration comes about on a typical rural settlement. We extracted the prototypical land-change patches from deforestation maps (Escada *et al.* 2005a) for the period from 1985 to 2000, with 30 m of spatial resolution and 3-year intervals. We trained the decision-tree classifier with 46 instances of land-change patches (figure 11).

The decision tree in figure 11 uses two metrics. The first metric is CIRCLE (circumscribing circle), which distinguishes the linear patches from the others

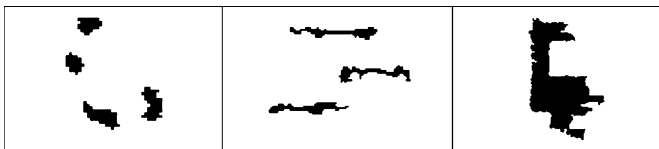


Figure 10. Spatial patterns in the Vale do Anari region: (from left to right) irregular, linear, regular.

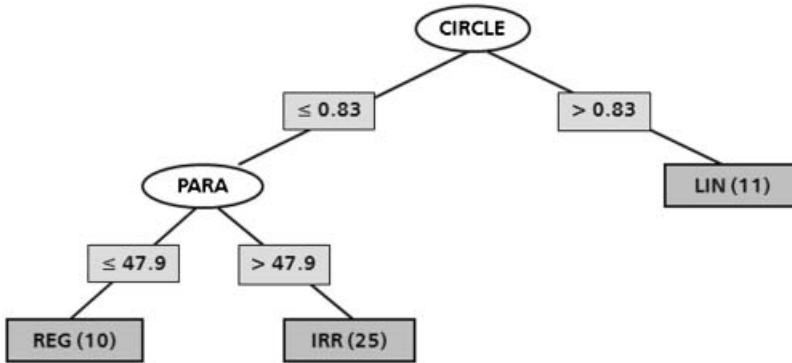


Figure 11. Decision tree for Vale do Anari spatial patterns. The metrics are: perimeter/area ratio (PARA) and circumscribing circle (CIRCLE).

(regular and irregular). Linear patches have a CIRCLE measure close to 1 (one). The PARA metric (perimeter-to-area ratio) distinguishes the *regular* patches from the *irregular* ones. Regular patches have a lower ratio than irregular patches. A leave-one-out cross-validation procedure showed a 98% accuracy rate.

Using the decision-tree classifier, we classified each map and then created a map showing the cumulative clearing patterns in Vale do Anari from 1985 to 2000 (figure 12). This map contains all new patches created by deforestation in the period, classified into one of the three types. We also created a temporal distribution, where we show the area for each patch type in a 3-year period (figure 13).

The cumulative map of clearings (figure 12) and the temporal distribution of patterns (figure 13) indicate a concentration of land ownership. In the earlier stages of the rural settlement, the dominant clearing patterns are linear and irregular. Linear patterns appear first, resulting from roadside clearings of household colonists. Then, as small-scale settlers deforest their parcels, we obtain irregular patterns. Linear and irregular clearing patterns match the land-use strategies of colonists in different occupation stages. Finally, as cattle ranchers buy the parcels from colonists, regular patterns start to appear on the fringes of irregular patterns.

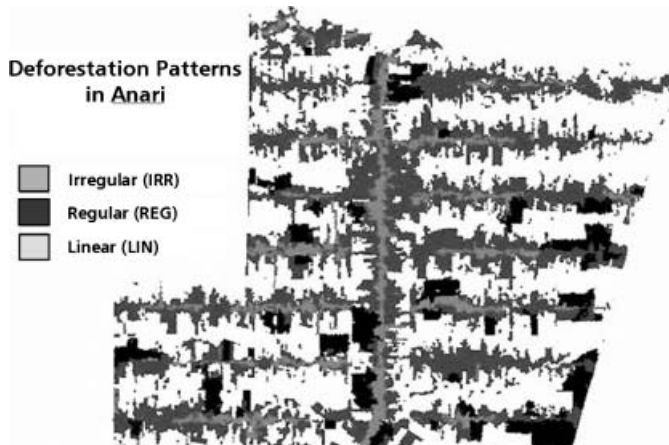


Figure 12. Cumulative deforestation patterns in Vale do Anari (1985–2000).

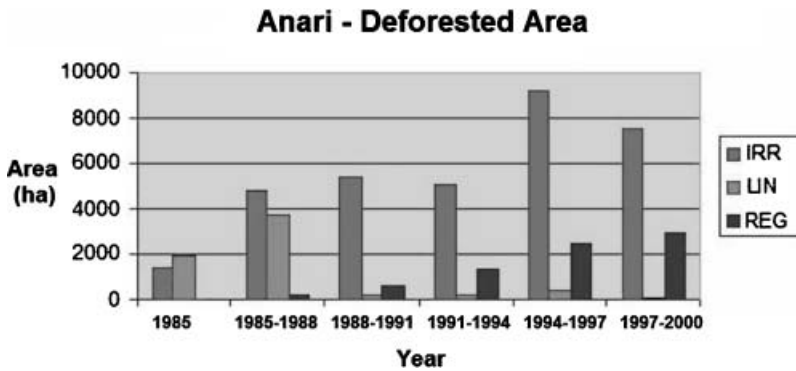


Figure 13. Distribution of deforestation patterns in Vale do Anari (1985–2000): irregular (IRR), linear (LIN), regular (REG).

From 1988 onwards, regular patterns grow progressively. They reach almost 30% of the deforestation in the period 1987–2000. This shows a marked land concentration, indicating that the government plan for settling many colonists in the area has been largely frustrated. The occurrence of land concentration in Vale do Anari was confirmed by fieldwork (Escada *et al.* 2005a). We identified and located 23 farms created by land concentration. We found that 87% of the classified land-change patches agreed with fieldwork data, showing that the data-mining method performs well.

5. Conclusions

This paper proposes a method for classifying land-change patches obtained from remote-sensing image databases, and as such the method addresses the problem of describing agents of land-use change. The method combines techniques from data mining, digital image processing, and landscape ecology. Pattern classification in maps extracted from classified images of distinct dates enables land-change patches to be associated with causative agents. The results from the case studies show that pattern-classification techniques associated with remote-sensing image interpretation are a step forward in understanding and modelling land-use change. The proposed method also enables a more effective use of the large land remote-sensing image databases available in agencies such as USGS, ESA, and INPE.

The proposed method points out that patch metrics can be used to identify agents of land-use change. Further experiments are necessary to improve the method, to test alternatives for image-segmentation algorithms and for patch classifiers. The limits of the current method include the two-dimensional nature of land-use maps. An extension of the method would be to combine spatial information (patch metrics) with spectral information (pixel and region trajectories in multitemporal images).

Future research directions in remote-sensing image mining include tracking individual trajectories of change. Patterns found in one map are linked to those in earlier and later maps, thus enabling a description of the trajectory of change in each land-change patch. The current method aggregates land-change patches of the same type. A more sophisticated approach would be to describe how each land-change patch evolves, including operations such as merging of adjacent regions. This description would allow the data mining to describe when two irregular areas of

land use (associated with small settlers) were merged. It would also show when the merged region was extended with a regular pattern (suggesting that a large cattle ranch had been established). This description could increase even more the ability to understand the land-use changes that are detectable in our remote-sensing image databases.

Acknowledgements

We would like to thank the anonymous reviewers of this paper for their useful comments. Gilberto Camara's work is partially funded by CNPq (grants PQ—300557/19996-5 and 550250/2005-0) and FAPESP (grant 04/11012-0). Marcelino Silva's work is supported by UERN and funded by CAPES. We thank Wilson Pagani from INCRA (Jaru) and Romain Taravella for the helpful information achieved during the fieldwork in Vale do Anari (Rondônia) and São Félix do Xingu (Pará) regions.

References

- ALMEIDA, A.L.O. and CAMPARI, J.S., 1996, *Sustainable Settlement in the Brazilian Amazon* (Washington, DC: World Bank Publications).
- ALVES, D., 2002, Space-time dynamics of deforestation in Brazilian Amazônia. *International Journal of Remote Sensing*, **23**, pp. 2903–2908.
- ALVES, D., ESCADA, M.I.S., PEREIRA, J.L.G. and LINHARES, C.D.A., 2003, Land use intensification and abandonment in Rondônia, Brazilian Amazônia. *International Journal of Remote Sensing*, **24**, pp. 899–903.
- AMARAL, S., MONTEIRO, A.M.V., CAMARA, G., ESCADA, M.I.S. and AGUIAR, A.P.D., 2007, Redes e conectividades na estruturação da frente de ocupação do Xingu-Iriri no Pará (Networks and connectivities in the structure of the occupation front in the Xingu-Iriri region of Pará). *Geografia*, **31**, pp. 655–675.
- ANDERSEN, L. and REIS, E., 1997, Deforestation, development, and government policy in the Brazilian Amazon: An econometric analysis (IPEA report 76-1997), IPEA. Available online at: <http://www.ipea.gov.br/pub/td/td0513.pdf> (accessed 29 March 2005).
- BATISTELLA, M., ROBESON, S. and MORAN, E., 2003, Settlement design, forest fragmentation, and landscape change in Rondonia, Amazonia. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 805–812.
- BECKER, B., 2004, *Amazônia—Geopolítica na Virada do III Milênio (Amazonia—Geopolitics on the Verge of the Third Millennium)* (Rio de Janeiro: Garamond).
- BINS, L., FONSECA, L. and ERTHAL, G., 1996, Satellite imagery segmentation: A region growing approach. In *VIII Brazilian Symposium on Remote Sensing*, E. Novo (Ed.), (São José dos Campos, Brazil: INPE), pp. 677–680.
- CÂMARA, G., EGENHOFER, M., FONSECA, F. and MONTEIRO, A.M., 2001, Whats in an image? In *Spatial Information Theory: Foundations of Geographic Information Science. International Conference, COSIT 2001*, D. Montello (Ed.), (Santa Barbara, CA: Springer), pp. 474–487.
- CÂMARA, G., SOUZA, R., FREITAS, U. and GARRIDO, J., 1996, SPRING: Integrating remote sensing and GIS with object-oriented data modelling. *Computers and Graphics*, **15**, pp. 13–22.
- CAMPARI, J.S., 2005, *The Economics of Deforestation in the Amazon: Dispelling the Myths* (Cheltenham, UK: Edward Elgar).
- DALE, V.H., O'NEILL, R.V., SOUTHWORTH, F. and PEDLOWSKI, M., 1994, Modeling effects of land management in the Brazilian Amazonian settlement of Rondonia. *Conservation Biology*, **8**, pp. 196–206.
- ESCADA, M.I.S., MONTEIRO, A.M., AGUIAR, A.P., CARNEIRO, T. and CAMARA, G., 2005a, Análise de padrões e processos de ocupação para a construção de modelos na

- Amazônia (Analysis of land use patterns and processes for the construction of models in Amazonia). In *XII Brazilian Symposium on Remote Sensing*, D. Valeriano (Ed.), (Goiania, Brazil: SELPER), pp. 2973–2983.
- ESCADA, M.I.S., VIEIRA, I.C.G., AMARAL, S., ARAÚJO, R., VEIGA, J.B.D., AGUIAR, A.P.D., VEIGA, I., OLIVEIRA, M., PEREIRA, J.L.G., FILHO, A.C., FEARNSIDE, P.M., VENTURIERI, A., CARRIELLO, F., THALES, M., CARNEIRO, T.S., MONTEIRO, A.M.V. and CÂMARA, G., 2005b, Padrões e processos de ocupação nas novas fronteiras da Amazônia: O interflúvio do Xingu/Iriri (Land use patterns and processes in Amazonian new frontiers: The Xingu/Iriri region). *Estudos Avançados*, **19**, pp. 9–23.
- EWERS, R.M. and LAURENCE, W.F., 2006, Scale-dependent patterns of deforestation in the Brazilian Amazon. *Environmental Conservation*, **33**, pp. 203–211.
- FEARNSIDE, P.M., 2002, Biodiversity as an environmental service in Brazil's Amazonian forests: Risks, value and conservation. *Environmental Conservation*, **26**, pp. 305–321.
- IMBERNON, J. and BRANTHOMME, A., 2001, Characterization of landscape patterns of deforestation in tropical rain forests. *International Journal of Remote Sensing*, **22**, pp. 1753–1765.
- INPE, 2005, Monitoramento da floresta amazônica brasileira por satélite (monitoring the Brazilian Amazon forest by satellite). INPE. Available online at: <http://www.obt.inpe.br/prodes/> (accessed 10 October 2005).
- KUHNERT, C., HELBING, D. and WEST, G.B., 2006, Scaling laws in urban supply networks. *Physica A*, **363**, pp. 96–103.
- LAMBIN, E.F., GEIST, H.J. and LEPELERS, E., 2003, Dynamics of land-use and land-cover change in tropical regions. *Annual Review of Environment and Resources*, **28**, pp. 205–241.
- MALANSON, G.P., ZENG, Y. and WALSH, S.J., 2006, Landscape frontiers, geography frontiers: lessons to be learned. *Professional Geographer*, **58**, pp. 383–396.
- MAS, J.F., 1999, Monitoring land-cover changes: A comparison of change detection techniques. *International Journal of Remote Sensing*, **20**, pp. 139–152.
- MCGARIGAL, K., 2002, Landscape pattern metrics. In *Encyclopedia of Environmentrics*, A.H. El-Shaarawi and W.W. Piegorsch (Eds), pp. 1135–1142 (Chichester, UK: Wiley).
- MCGARIGAL, K. and MARKS, B., 1995, Fragstats: Spatial pattern analysis program for quantifying landscape structure. Available online at: www.umass.edu/landeco/research/fragstats/fragstats.html (accessed 27 October 2005).
- MEINEL, G. and NEUBERT, M., 2004, A comparison of segmentation programs for high resolution remote sensing data. *International Archives of Photogrammetry and Remote Sensing*, **XXXV**, pp. 1097–1105.
- MERTENS, B. and LAMBIN, E., 1997, Spatial modeling of deforestation in southern Cameroon: Spatial disaggregation of diverse deforestation processes. *Applied Geography*, **17**, pp. 143–162.
- OLIVEIRA FILHO, F.B. and METZGER, J.P., 2006, Thresholds in landscape structure for three common deforestation patterns in the Brazilian Amazon. *Landscape Ecology*, **21**, pp. 1061–1073.
- PERALTA, P. and MATHER, P.M., 2000, An analysis of deforestation patterns in the extractive reserves of Acre, Amazonia from satellite imagery: A landscape ecological approach. *International Journal of Remote Sensing*, **21**, pp. 2555–2570.
- PERZ, S.G. and SKOLE, D.L., 2003, Social determinants of secondary forests in the Brazilian Amazon. *Social Science Research*, **32**, pp. 25–60.
- PPFAFF, A., 1999, What drives deforestation in the Brazilian Amazon? Evidence from satellite and socio-economic data. *Journal of Environmental Economics and Management*, **37**, pp. 26–43.
- QUINLAN, R., 1993, *C4.5: Programs for Machine Learning* (San Francisco: Morgan Kaufmann).
- SCHIMINK, M. and WOOD, C.H., 1992, *Contested Frontiers in Amazonia* (New York: Columbia University Press).

- SHIMABUKURO, Y., BATISTA, G., MELLO, E., MOREIRA, J. and DUARTE, V., 1998, Using shade fraction image segmentation to evaluate deforestation in Landsat thematic mapper images of the Amazon region. *International Journal of Remote Sensing*, **19**, pp. 535–541.
- SHUKLA, J., NOBRE, C. and SELLERS, P., 1990, Amazon deforestation and climate change. *Science*, **247**, pp. 1322–1325.
- SILVA, M.P.S., CÂMARA, G., SOUZA, R.C.M., VALERIANO, D.M. and ESCADA, M.I.S., 2005, Mining patterns of change in remote sensing image databases. In *The Fifth IEEE International Conference on Data Mining*, J. Han and B. Wah (Eds), (Houston, TX: IEEE).
- SOUTHWORTH, J., NAGENDRA, H. and TUCKER, C., 2002, Fragmentation of a landscape: Incorporating landscape metrics into satellite analyses of land-cover change. *Landscape Research*, **27**, pp. 253–269.
- TURNER, M.G., 1989, Landscape ecology: The effect of pattern on process. *Annual Review of Ecology and Systematics*, **20**, pp. 171–197.
- VAROTSOS, C., 2005a, Power-law correlations in column ozone over Antarctica. *International Journal of Remote Sensing*, **26**, pp. 3333–3342.
- VAROTSOS, C., 2005b, Modern computational techniques for environmental data; application to the global ozone layer. *Lecture Notes in Computer Science*, **3516**, pp. 504–510.
- WITTEN, I.H. and FRANK, H., 1999, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations* (San Francisco: Morgan Kaufmann).