Stratifying land use/land cover for spatial analysis of disease ecology and risk: an example using object-based classification techniques

David E. Koch, Rhett L. Mohler, Douglas G. Goodin

Department of Geography, Kansas State University, Manhattan, KS 66506-2904, USA

Abstract. Landscape epidemiology has made significant strides recently, driven in part by increasing availability of land cover data derived from remotely-sensed imagery. Using an example from a study of land cover effects on hantavirus dynamics at an Atlantic Forest site in eastern Paraguay, we demonstrate how automated classification methods can be used to stratify remotely-sensed land cover for studies of infectious disease dynamics. For this application, it was necessary to develop a scheme that could yield both land cover and land use data from the same classification. Hypothesizing that automated discrimination between classes would be more accurate using an object-based method compared to a per-pixel method, we used a single Landsat Enhanced Thematic Mapper+ (ETM+) image to classify land cover into eight classes using both per-pixel and object-based classification algorithms. Our results show that the object-based method achieves 84% overall accuracy, compared to only 43% using the per-pixel method. Producer's and user's accuracies for the object-based map were higher for every class compared to the per-pixel classification. The Kappa statistic was also significantly higher for the object-based classification. These results show the importance of using image information from domains beyond the spectral domain, and also illustrate the importance of object-based techniques for remote sensing applications in epidemiological studies.

Keywords: landscape epidemiology, land use, land cover, remote sensing, object-based classification, hantavirus, Paraguay.

Introduction

Landscape epidemiology is grounded in the notion that disease occurs in spatial patterns arising from underlying variation in environmental conditions (physical or biological) that can be delimited on maps (Pavlovsky, 1966; Ostfeld et al., 2005). Geospatial analysis technologies such as remote sensing (RS) and geographical information systems (GIS) provide tools for gathering and analyzing these spatial data across a wide spectrum of spatial scales. RS has been applied to a variety of landscape epidemiological studies

(Beck et al., 2000), and has been especially effective for analysis of a number of vector-borne and zoonotic diseases with environmental co-factors. Some of these diseases include malaria (Mushinzimana et al., 2006), Lyme disease (Brownstein et al., 2005), Chagas disease (Kitron et al., 2006), West Nile fever (Rogers et al., 2002), hantavirus pulmonary syndrome (Glass et al., 2000), Ebola Hemorrhagic fever (Pinzon et al., 2004), and Rift Valley fever (Linthicum et al., 1987; Martin et al., 2007). Curran et al. (2000) noted that one of the principal advantages of satellite RS for analyzing disease-related environmental factors was the capability for rapid and repetitive collection of information, even from remote or inaccessible places. This capability for generating a dense, repetitive spatio-temporal data stream has led a number of researchers to explore the use of satellite imagery as a fundamental data source for disease risk maps, which

Corresponding author:
Douglas G. Goodin
Department of Geography, Kansas State University
118 Seaton Hall
Manhattan, KS 66506-2904, USA
Tel. +1 785 532 3411; Fax +1 785 532 7310
E-mail: dgoodin@ksu.edu

incorporate vector distribution and landscape elements into a spatial model of disease occurrence probability (Glass et al., 2006). Although these forecast models are still being developed, they offer promise as tools for mitigating environmental health risks.

Retrieval of land use/land cover information is a common application of RS. Within landscape epidemiology, this type of information is frequently used to identify and map vector or host habitat (Kazmi and Usery, 2001). Most classification schemes are based on clustering of individual pixels in spectral feature space, however this approach ignores the spatial and geometric information inherent in imagery, which can be useful for correctly identifying cover type. Object-based classification techniques (Jansen and van Amsterdam, 1991) provide an alternative means of mapping land cover or habitat type useful for studying landscape influences on disease process. These techniques differ from per-pixel approaches in that they first segment the image into homogeneous regions ("objects"), which then become the units of analysis and classification. Object-based methods are sensitive to both the spectral and geometric properties of the imagery, and are thus appropriate methods for use when mapping land cover classes where the discriminating characteristics of the class may include the shape and size of homogenous polygons within the scene (Jansen and van Amsterdam, 1991; Lobo et al., 1996).

In this paper, we present an example of object-based classification applied to deriving spatial information for disease ecology and risk mapping.

Object-based classification example: hantavirus in Paraguay

Hantaviruses are zoonotic RNA viruses hosted by rodents of the Sigmadontinae subfamily. There are numerous subspecies or strains of hantavirus, each associated with a particular rodent host species. In the western hemisphere, the virus causes hantavirus pulmonary syndrome (HPS), an acute cardiopulmonary disease fatal in about 30-50% of all cases

(Schmaljohn and Hjelle, 1997). Since emerging in Paraguay in 1995 (Williams et al., 1997), hantaviruses have been shown to be endemic throughout Paraguay, including the eastern Atlantic Forest area where this research was conducted (Chu et al., 2006). Managing the impacts of HPS outbreaks can be a significant public health problem, particularly in a less-developed country like Paraguay that lacks extensive medical infrastructure. Better understanding of the ecology of hantavirus and the rodents that carry it could lead to improved surveillance efforts and ultimately to prevention of the disease.

Risk of HPS infection in a human population is dependent on the probability of encountering transmissible virus from an infected host. From a landscape perspective, this is determined by the distribution of the host rodent population and the human population on the landscape. Both of these depend to some degree on land use and land cover, although the exact nature of the relationship between host, landscape, and human population varies. The analysis presented here is a component of a multidisciplinary investigation of the role of land cover in the ecology of hantavirus dynamics. Addressing the major hypothesis of this research project, that human land cover disturbance is associated with increased occurrence of hantavirus in both rodents and humans (Goodin et al., 2006), presents challenges to the use of RS for mapping landscape epidemiological factors, since both land cover and land use must be extracted from a single classification. Foremost among these challenges is separation of land clearing associated with various forms of agriculture.

Agriculture (including pasturing) in the study site typically comes in two forms, one associated with large-scale commercial activity, the other characteristic of smaller scale production by small land holders. Differentiating large from small-scale agriculture using image data is challenging. Typically, large-scale agriculture results in very large cleared fields (100s - 1000s of ha), often with regular, rectangular borders (Fig. 1). In contrast, small-holder agriculturists typically clear parcels of 10-20 ha on

which they practice a mixture of subsistence and mar-
ket crop production. These parcels are often located
in groups along roads with some remnant forest sep-
arating them, resulting in the characteristic "piano
key" or "fishbone" deforestation pattern observed
throughout the tropics (see Fig. 1; Skole and Tucker,
1993). In areas where fields are longer established,
several smallholdings often coalesce into a larger
cleared area, resembling large-holder clearing.

The importance of separating the two types lies in
the very different settlement and population pat-
terns associated with each. Typically, large-scale
agricultural parcels are not densely populated,
whereas each smallholder parcel is usually occupied
by a tenant family, often of 10 or 12 related persons.
Smallholder agricultural areas also contain settle-
ments consisting of concentrations of dwellings with
a few businesses interspersed. Thus, concentrated
areas of smallholdings are relatively densely popu-
lated, compared to their surrounding land uses.

Both the large and small-scale agricultural cover
types are spectrally similar to one another, in fact
they are essentially the same land cover, although
not the same land use. In the study area, these two
agricultural land use types also tend to be inter-
spersed, producing a complex surface mosaic where
discriminating characteristics are frequently geomet-
ric, rather than spectral (see Fig. 1). The complex
nature of this surface cover mosaic suggests the use
of an object-based classification approach for con-
structing a map suitable for analyzing the spatial
relationships between the hantavirus host popula-
tion and the human population.

Materials and methods

Study area

This research was conducted at a site within the
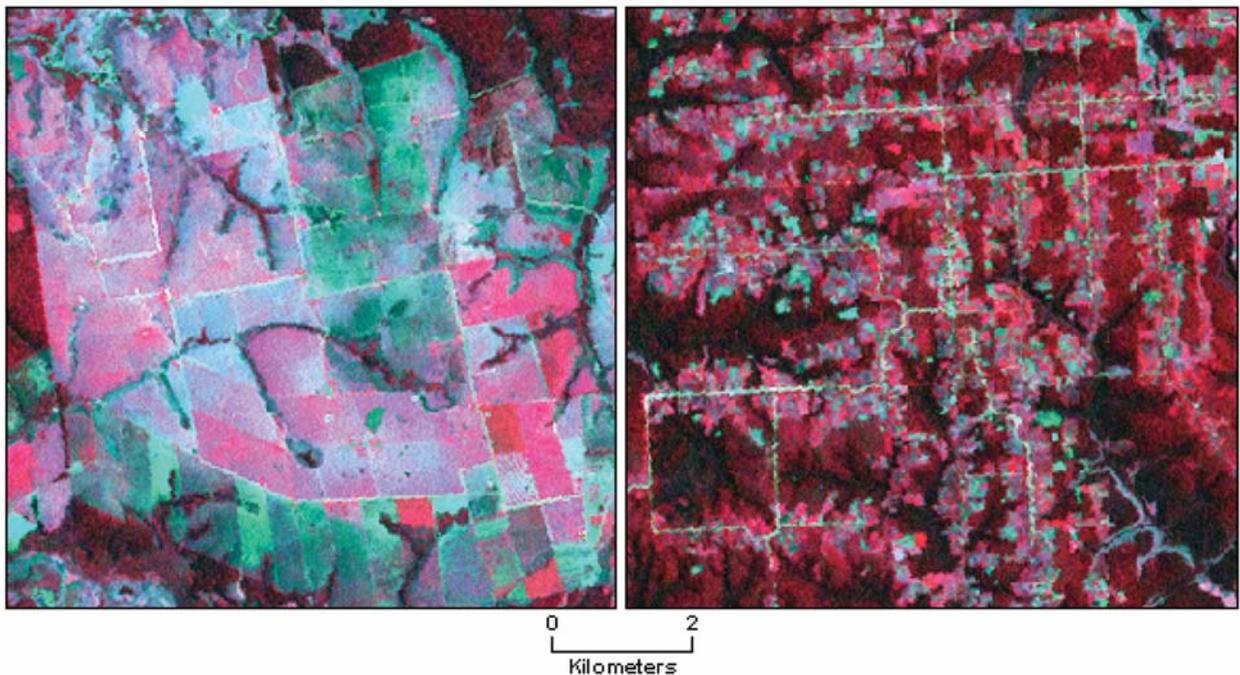Reserva Natural Bosque Mbaracayú (RNBM), an

Fig. 1. Portion of an ETM+ false color composite (Bands 4, 3, 2) showing large-scale agricultural clearance (left panel) and
small-scale agricultural clearing (right panel). Both of these land uses have similar land cover, hence similar spectral reflectance.
The two types can be discriminated by their geometric properties. Note the characteristic "fish-bone" clearing pattern associ-
ated with small-scale agriculture.

ecological reserve located within the Paranã-Paraíba subtype of the Interior Atlantic Forest (Fig. 2). Although the RNBM is a biosphere reserve, it is not all protected land. Within the reserve there are a number of inhabited areas, included indigenous settlements (Ache, Guarani), rural Paraguayan communities, and immigrants (mostly from Brazil). The reserve contains an inner core dominated by one large forest fragment (the largest extant Interior Atlantic Forest fragment in Paraguay, see Fig. 2). Forest conversion is illegal in this area. Throughout the rest of the study area, active forest conversion occurs, typically near the various inhabited areas. This conversion is primarily of forested land to agriculture (crop or pasture), interspersed by smaller

Fig. 2. The study site in the Interior Atlantic forest, eastern Paraguay.

forest remnants. The study area also contains large areas of cerrado, natural savannas maintained by fire (dry cerrado) or water (wet cerrado). Cerrados are characterized by grasses, woody shrubs and palms. Some cerrado has been converted to pasture or cropland (Sarmiento, 1983).

Data

Classifications were performed on a Landsat ETM+ image acquired on 28 February 2003 (path 225/row 77). This image was collected prior to the failure of the ETM+ scan line corrector (Howard and Lacasse, 2004). Data were prepared for analysis by subsetting the study area from the full scene and georectifying it to the Universal Transverse Mercator (UTM) grid (zone 21S) using 1st-order polynomial (Richards and Jia, 1999) with control points derived from in-situ survey and examination of Quickbird high resolution satellite imagery. Total RMS error for the georectification was less than 1.0. To insure that the Quickbird imagery was itself spatially accurate, a number of points from these images were compared to ground points with known coordinates. There was close agreement between these comparison points. Nearest neighbour re-sampling was used to estimate new pixel brightness values. Following the recommendation of Song et al. (2001) raw pixel digital numbers were corrected for atmospheric backscatter using the IDOS model (Chavez, 1988), assuming an Ångstrom exponent of 2 (Iqbal, 1983).

Classification scheme

Because of the specialized nature of this classification, it was necessary to develop a classification scheme where the classes included aspects of both land use and land cover. Classification categories needed to convey meaningful information about rodent habitat, while at the same time being interpretable in terms of human land use and the presence of human inhabitants. We therefore adopted a hybrid classification scheme in which cover classes were

defined in terms of both land use and land cover. The eight classes in the classification scheme were determined from a number of sources including previous land cover classifications (Naidoo and Hill, 2006), on-site observations, and expert opinions from researchers and others familiar with the study area.

Our classification lacks some of the detail of other classifications of this area (e.g. Burgos and Rodas, 2001; Naidoo and Hill, 2006). Unlike these classifications, we chose to represent the forest by only two classes, labeled "Forest" and "Disturbed forest". These simplified forest categories better suited the known habitat preferences of hantavirus host rodents in the area, especially *Akodon montensis* and *Oligoryzomys fornesi* (Nitikman and Mares, 1987; Emmons and Feer, 1990; Gentile and D'Andrea, 2000; Figueiredo and Fernandez, 2004). Non-forest classes were defined by a combination of disturbance history (selective logging) and vegetation. Cerrado was divided into two classes, wet and dry. Wet cerrado tends to occur in low-lying and riparian areas with persistently wet soil. It is characterized by natural open (non-forested) cover dominated by a mix of native and invasive/introduced grasses and woody plants. Dry cerrado is found in areas where fire is used as a management tool. Typical vegetation on the dry cerrado included woody plants, palms, and fire-tolerant grasses and forbs. Pasture areas were also dominated by native and introduced grasses, but in areas cleared of their original cover. Dry pasture was found in cleared forest areas. Wet pasture consisted of wet cerrado managed for cattle. Agricultural areas are dominated by a variety of cultivated crops, generally grown on cleared forest land. Crops vary depending on cultivation type. Market crops dominate on large holdings, while a diverse mix of subsistence and market crops are grown in smallholder parcels. The classes are summarized in Table 1.

Classification methodology

The goal of this study was to evaluate object-based classification strategies for extracting the land cover

classes described in Table 1. We performed this evaluation by comparing object-based against per-pixel classification. We chose one approach for each of the two classification strategies. For the per-pixel classification, we used a supervised, maximum likelihood classification algorithm (Richards and Jia, 1999), implemented in the ENVI software package, version 4.2. The maximum likelihood algorithm works by developing a set of membership probability functions for each class using training data selected from the imagery. These membership probabilities are based on proximity between classes in *N*-dimensional feature space, where *N* is the number of spectral features used in the classification. Because only spectral information is used to develop the discriminant functions, the maximum likelihood algorithm is restricted to using spectral reflectance information as criteria for discriminating between classes. Training site selection was guided by ground truth obtained from two visits to the study site (October 2002 and June 2005), together with inspection of high resolution satellite data. Training data were gathered from a number of sites for each cover class. To avoid mixed training pixels, we were careful to avoid transitional or edge areas when selecting samples from the training sites. Because the maximum likelihood algorithm is sensitive to non-normally distributed training data, we selected a minimum of 800 training pixels for each cover category. Prior to classification, we used the transformed divergence index (TDI, see Jensen, 1995) to determine the optimum spectral features for classification. Based on the results of TDI, we used ETM+ bands 1, 3, 4 and 6 to classify the image. Results of this initial classification were improved by application of a 3 x 3 majority filter to reduce or eliminate random pixel classification errors and noise (Gurney and Townshend, 1983). The resulting classification map is shown in Figure 3.

Our implementation of object-based classification was a two-step process. In the first or segmentation step, spatial objects were formed. Objects are defined as groups of adjacent pixels treated as a single entity (Hay et al., 2001). Each object is composed of similar pixel values and possesses intrinsic

D.E. Koch et al. - Geospatial Health 2(1), 2007, pp. 15-28

Table 1. Description of land use/land cover categories.

Class	Dominant land cover	Primary land use	Characteristics
Forest	Trees from the Leguminosae, Lauraceae and Myrtaceae families	Indigenous hunting/gathering†	Dense forest with little or no evidence of human disturbance. No permanent habitation.
Forest-disturbed	Trees from the Leguminosae, Lauraceae and Myrtaceae families	Logging, indigenous and Paraguayan farmer hunting/gathering†	Forest with visible evidence of human disturbance, often selective logging. Disturbance is not necessarily recent. Often characterized by more open canopy and more dense understory. No permanent habitation.
Cerrado-wet	*Andropogon* spp., *Xyris jupucai, Rhynchospora* spp., *Axonopus* spp.	Large animal grazing	Natural clearing. Occurs along drainages, riparian areas and forest edges with persistently wet soils. Little habitation.
Cerrado-dry	*Butia paraguayensis, Rhynchospora* spp., *Paspalum plicatum, Andropogon* spp.	Large animal grazing	Natural clearing, but occurs on drier soils. Little habitation.
Agricultural-smallholder	Mixed market and subsistence crops. *Manihot esculenta, Zea mays, Gossypium* spp., *Citrus* spp., *Andropogon* spp., *Paspalum* spp.	Subsistence and small-scale commercial agriculture. May include some small/large animal grazing	Mixed crop agriculture characterized by small, cultivated parcels (<20 ha is typical), usually cleared from forest. Often interdigitated with forest remnant, producing a 'piano-key' appearance. Typically occupied by land owner family, resulting in higher population density.
Agricultural-largeholder	Market crops. *Glycine max, Triticum* spp.	Commercial agriculture	Monoculture cropland. Crops can vary with season. Typically very large, regularly shaped fields formed by clearing forest.
Pasture-wet	*Xyris jupicai, Rhynchospora* spp., *Andropogon* spp., *Paspalum* spp.	Livestock grazing	Large, regularly shaped parcels located on wetter soils or in riparian areas.
Pasture-dry	*Paspalum* spp., *Andropogon* spp., *Rhynchospora* spp.	Livestock grazing	Large, regularly shaped parcels cleared from forest, located on drier soils.

†=see Naidoo and Hill, 2006

size, shape and relationship to the overall image topology (Yu et al., 2006). Objects were segmented using the fractal net evolution approach (FNEA) as implemented in the eCognition software, v4.2 (Baatz et al., 2004). FNEA uses a bottom-up method for building scene objects, where each object begins with a single pixel. Objects are merged pair wise, using a merging criteria that minimizes size-weighted object heterogeneity, where heterogeneity is defined in terms of both object spectral and geometric properties (Benz et al., 2004). The segmentation algorithm depends on a set of user-defined parameters that control the scale, shape factor and compactness of the resulting objects. The

values that yielded an optimal segmentation were: a scale parameter of 16, shape factor of 0.1 and compactness/smoothness ratio of 0.5/0.5.

In the second step, objects were sorted into classes. Each object in the segmented image is described by a number of spectral, geometric and textural features, any or all of which can be used in the classification. We used the mean object spectral value and standard deviations for ETM+ bands 1-5 and 7 (minus the thermal band) to capture the magnitude and heterogeneity of reflectance in the data.

In addition, we used the dissimilarity and standard deviation values derived from the grey-level co-occurrence matrix (GLCM) for each input band as
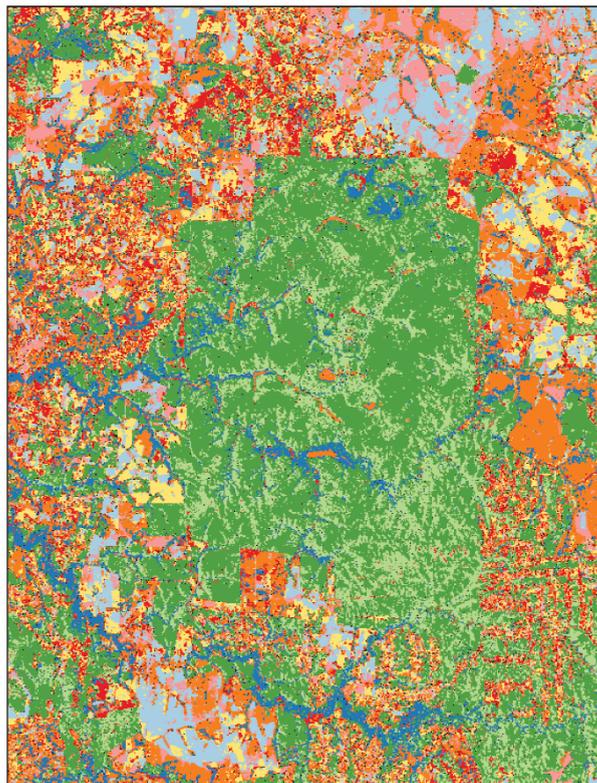

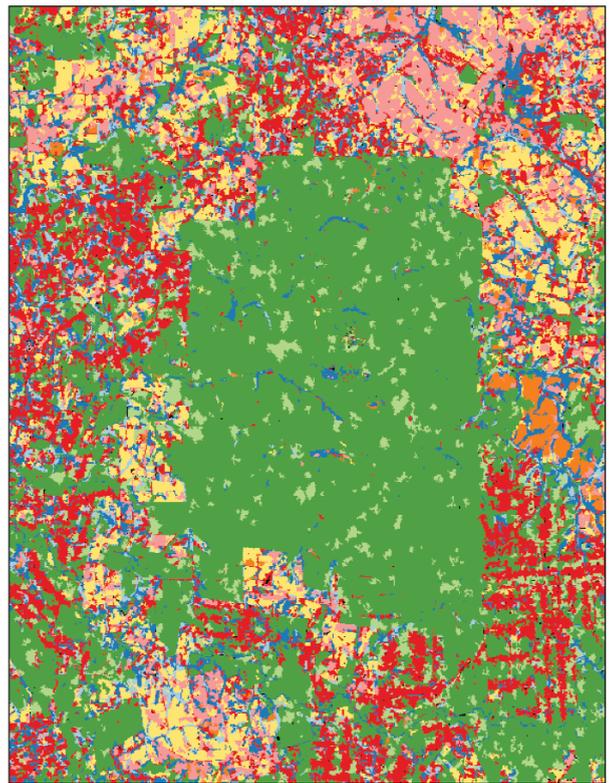Fig. 3. Classified map of the study site derived from a per-pixel algorithm.

Fig. 4. Classified map of the study site derived from a object-based algorithm.

metrics for object texture (Haralick and Shanmugam, 1974). Object shapes were characterized using the area and asymmetry metrics (Baatz et al., 2004). These metrics were submitted to a minimum distance decision algorithm with fuzzy class boundaries. Class boundaries were hardened by assignment of the object to the most probable class, except in cases where one or more classes were nearly equally likely. In these cases, a proximity algorithm was used to harden the classification based on contextual criteria.

Accuracy assessment

Accuracy assessment was done in similar ways for each of the two classifications. Validation sites were selected by proportional stratified random sampling from the classified image (Beyer, 2004). For the per-pixel classification, validation pixels were selected randomly from within the area of the objects selected for validating the object-based classification. Sites in which the cover type could not be identified by either field data or high resolution imagery were eliminated, resulting in a sample of 632 validation pixels from the per-pixel classification, and 227 objects from the object-based map. For the object-based classification, accuracy assessment was performed on a per-object basis. Overall numbers of validation samples in the object-based classification were smaller, since segmentation results in fewer objects within the scene.

Validation objects and pixels were evaluated for correctness, and the results were tabulated into error matrices for each of the classifications (Tables 2 and 3). A number of quantitative accuracy metrics were computed from these error matrices. These metrics included total accuracy, the ratio of correctly classified pixels to total validation samples, as well as the producer (Type I) and user (Type II) accuracies for each cover class (Jensen, 1995). In addition, we calculated Cohen's Kappa statistic (Cohen, 1960; Congalton and Mead, 1983), an index of improvement of the classification compared to random pixel assignment. As an accuracy

metric, Cohen's Kappa has the added benefit that two values can be compared for statistical difference using a test based in the z-statistic (Congalton and Mead, 1983).

Results and discussion

Comparison of the classification show the clear superiority of the object-based method for the current application. Examination of the classified maps (Figs. 3 and 4) show similar overall pattern, but the map resulting from per-pixel classification is much noisier and the regions associated with the various classes are poorly defined. Quantitative results confirm these observations (Tables 2 and 3). The object-based classification correctly classified 84% of scene objects into the correct class, whereas only 43% of pixels in the per-pixel classification were correct. Quantitative comparison confirms that the Kappa statistic for the object-based classification is significantly higher than that of the per-pixel classification (z = 58.6, p <0.001). Despite the obvious difference between the two methods, it is instructive to look at the classifications in more detail.

Both classification strategies showed the greatest agreement in the forest classes, particularly in the Forest class. User's accuracy (equivalent to Type I statistical error) is identical for the "Forest" categories in the two classifications, with 82% of pixels or objects correctly classified. This agreement suggests that discriminating this class of forest cover must depend primarily on spectral criteria, since the per-pixel classification is insensitive to any other aspect of the data. Although the User's accuracy for the two classification maps is the same, there is still a valid argument for preferring the object-based classification. In the object-based classification, all of the misclassified objects were from the "Disturbed forest" class. In the pixel-based classification, only 6 of the 16 confused pixels were from the similar "Disturbed forest" class. The other 10 misclassified pixels were from a less similar agricultural class. Producer's accuracy (i.e. Type II statistical error) for the "Forest" class was also somewhat better for the object-based

Table 2. Accuracy assessment for per-pixel based classification.

Map class	Reference class								Total	User's accuracy
	Forest	Forest-disturbed	Cerrado-wet	Cerrado-dry	Agricultural-largeholder	Agricultural-smallholder	Pasture-wet	Pasture-dry		
Forest	71	6			10					0.82
Forest-disturbed	8	33	9		37	12			87	0.42
Cerrado-wet	4		15		20	2	20		79	0.25
Cerrado-dry			5	16	27	7		14	61	0.21
Agricultural largeholder	10	6	4	4	64	19	2	1	75	0.66
Agricultural-smallholder	8	4	4	9	40	18	3	1	128	0.28
Pasture-wet	1		5		21	25	3	1	90	0.06
Pasture-dry			4	15	11	10	1	22	49	0.35
Total	102	49	46	44	230	93	29	39	64	
Producer's accuracy	0.70	0.67	0.33	0.36	0.37	0.27	0.10	0.56		

Overall accuracy = 0.43

Kappa = 0.32

Table 3. Accuracy assessment for object-based classification.

Map Class	Reference class								Total	User's accuracy
	Forest	Forest-disturbed	Cerrado-wet	Cerrado-dry	Agricultural-largeholder	Agricultural-smallholder	Pasture-wet	Pasture-dry		
Forest	27	6							33	0.82
Forest-disturbed	4	28							32	0.88
Cerrado-wet			11			1	2		14	0.79
Cerrado-dry			1	20	2			4	27	0.74
Agricultural largeholder				2	42	3	1	1	49	0.86
Agricultural-smallholder						17		1	18	0.94
Pasture-wet	1					1	20	1	23	0.87
Pasture-dry				4	1	1	1	22	29	0.81
Total	32	35	12	33	46	24	23	29		
Producer's accuracy	0.84	0.80	0.92	0.91	0.91	0.71	0.87	0.76		

Overall accuracy = 0.84

Kappa = 0.81

classification, and once again the confusion was almost entirely between objects from the forest categories. In contrast, only 70% of "Forest" pixels in the image were classified as such by the maximum likelihood algorithm, and the remaining 30% of pixels were confused with a number of classes.

The "Disturbed forest" category was even more accurately classified by the object-oriented method, compared to the per-pixel approach. The low User's accuracy for the per-pixel classification (42%) resulted mainly from a large number of agricultural pixels (both largeholder and smallholder) being misclassified as "Disturbed forest". In contrast, the corresponding User's accuracy of the object-based classification is 88%, with all of the confused pixels belonging to the "Forest" class. Spectral confusion between "Disturbed forest" and some agricultural pixels is not as surprising as it may appear at first glance. Much of the agricultural land in the study site is recently converted forest, and often the boundaries between forest and agricultural field are indistinct, especially at the resolution of ETM+ data. In addition, the crop grown on the larger commercial fields is often manioc, a tuberous crop whose aboveground components are woody shrubs spectrally similar to some secondary forest growth in the "Disturbed forest" areas. Given the limitation on the available spectral feature space inherent in ETM+ bands, it is not surprising that these dissimilar land cover/land use classes might be spectrally confused.

Unlike all other non-forest classes, the two Cerrado cover types are not the result of deforestation. They are spectrally similar to the pasture and agricultural classes, but lack the regular geometric edges characteristic of deforestation and are heterogeneous at different spatial scales than pasture. These differences in heterogeneity and geometry can be incorporated into an object-based classification rule. In fact, the object-oriented method substantially outperformed the per-pixel method in classifying both the "Wet" and "Dry cerrado" types. In terms of User's accuracy, the maximum likelihood decision rule was unable to discriminate between "Dry cerrado" and "Dry pasture" or either of the agricultural classes. The "Wet cerrado" class also showed considerable confusion with Largescale agriculture and "Wet pasture". There was comparatively little confusion between the two cerrado types. Producer's accuracies for the Cerrado classes were slightly better, although still low. "Wet cerrado" showed more interclass confusion in Producer's accuracy than it did in User's, with confused pixels in every category but "Closed forest". "Dry cerrado" was mostly misclassified as "Dry pasture", but also showed confusion with the agricultural classes.

In terms of User's accuracy, the pattern of categorical confusion in the object-based classification of cerrado was similar to that of the per-pixel classification, although the overall accuracy was much higher. Like the per-pixel classification, most of the error associated with the cerrado classes resulted from confusion with the agricultural and pasture classes. The Producer's accuracies for the two cerrado classes were the highest of any of the cover classes. Error in producer's accuracy for "Wet cerrado" was entirely accounted for by a single object misclassified as "Dry cerrado". "Wet cerrado" showed some Producer's error due to confusion with pasture and agriculture, but the overall accuracies still exceeded 90%.

The maximum likelihood method showed almost no ability to discriminate "Wet pasture" pixels from most other types. Only three of the reference pixels classified as "Wet pasture" were placed in the correct category, resulting in an User's accuracy of only 6%. Producer's accuracy was only 10%, with most of the inaccuracy resulting from confusion with the "Wet cerrado" class. In contrast, the User's and Producer's accuracies for the object-oriented classification of "Wet pasture" were both 87%. The marked difference in results between the two classification strategies for "Wet pasture" illustrates the main advantage of the object-based methods for discriminating classes in a landscape like the one considered here. "Dry pasture" was also more accurately classified using the object-based method. Per-pixel classification resulted in a map where "Dry pasture" was fre-

quently confused with agriculture land use and "Dry cerrado". This confusion results from the spectral similarity of these cover types. As was the case with "Wet pasture", the object-based classification strategy accounted for the shape and geometry of the "Dry pasture" polygons, and was therefore able to classify them with much greater accuracy.

Pasture land is defined here as resulting from human disturbance (see Table 1), created primarily by clearing forest. This clearing results in a geometry characterized by straight, abrupt edges along boundaries. The segmentation process used here is sensitive to this geometry, resulting in objects whose shape is diagnostic of pasture land. In addition, the ability of the fuzzy decision rule to take into account shape-based variables results in a much more powerful set of discriminant rules, which is reflected in higher classification accuracies.

Accurate classification of the agriculture categories is particularly important to the research project for which this classification is intended. This is especially true for the small-scale agriculture class, because the majority of the human population in the study area is found within this class. "Smallholder agriculture" classification using the per-pixel approach was almost equally inaccurate from both the Producer's and User's perspective. An indicator of the inability of the maximum likelihood algorithm to discriminate between these classes is the fact that small-scale agriculture was, to some degree, confused with every other class. The confusion was greatest between large and small-scale agriculture, which is not surprising given the spectral similarity of the vegetation cover in the fields, but the confusion with other classes once again illustrates the basic shortcoming of over reliance on clustering in spectral feature space to differentiate classes. "Smallholder agriculture" is fairly easy to visually distinguish on the imagery (see Fig. 1), due mainly to shape and texture clues resulting from the juxtaposition of small polygons of cleared and uncleared land. The spatial intensity and small-scale of these parcels suggests that they are represented on the image by many mixed pixels, which would tend to plot into the margins of the class clouds in spectral feature space, confusing even a powerful classification algorithm like maximum likelihood. The additional information provided by the shape and texture of spatial objects allowed the automated classification algorithm to use the sort of information that a human classifier would use to discriminate these classes.

For the per-pixel classification, User's accuracy for the large-scale agriculture category is second only to the "Forest" class. Most of the confusion is with small-scale agriculture, as expected given the spectral similarities of the crops. Large-scale agriculture did not fare as well in Producer's accuracy, where it showed some confusion with every class. Although the per-pixel classification strategy was marginally more successful in discriminating large-scale agriculture pixels compared to the other classes, it was much less successful that the object based technique, for many of the same reasons already discussed with respect to other anthropogenic classes. The ability of the object-based method to use geometrical as well as spectral information to separate classes was the key to its superior performance.

Conclusions

The results reported here clearly show the superiority of the object-oriented technique for this land use/land cover application. Despite basing the entire classification on a single image, we were able to achieve an overall classification accuracy of 84%, with some individual classes being classified even more accurately. This value is comparable to other classifications for this region (Huang et al., 2007), which were achieved using integrative per-pixel methods and manual interpretation applied to multidate imagery and a simpler classification scheme. Our classification was achieved from a single image with all classes determined from one application of the classification algorithm. Some experimentation was necessary to determine the optimal parameters for the object-based classification, but successive classifications could use this same set of parameters,

simplifying the process of repeating this classification. It is likely that results could be improved even further by applying them to multidate imagery.

Use of satellite-derived land cover maps is an increasingly important component of landscape epidemiological studies of infectious disease. As we have illustrated here, mapping land cover for disease analysis often require conceptualizing land use and land cover in an unusual way, which can complicate the process of extracting useful information from satellite data. Object-based classification methods have certainly proved useful in this particular application, and might be well-applied to other, similar problems in land cover mapping. These methods are an increasingly important component of the 'toolkit' of any researcher using RS data. Application of these advanced classification algorithms in disease or health-related studies is a potential area of collaboration between epidemiologists, disease ecologists and RS scientists.

Acknowledgements

References

Baatz M, Benz U, Deghani S, Heynen S, Holtje A, Hofmann P, Ingenfelder I, Mimler M, Sohlbach M, Weber M, Wilhauck G, 2004. Ecognition User's Guide, v.4.0. Definiens, AG, Munich, Germany.

Beck LR, Lobitz BM, Wood BL, 2000. Remote sensing and human health: new sensors and new opportunities. Emerg Infect Dis 6, 217-227.

Benz UP, Hoffman P, Willhauck G, Lingenfelder I, Heynen M, 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J Photogram Rem Sens 58, 239-258.

Beyer HL, 2004. Hawth's analysis tools for Arc-GIS.

Brownstein JC, Skelly DK, Holford TR, Fish D, 2005. Forest fragmentation predicts local scale heterogeniety of Lyme disease risk. Oecologia 146, 469-475.

Burgos SM, Rodas O, 2001. Relaciones suelo-paisaja y mapeamiento geomorfico de la Reserva del Bosque Mbaracayu, Paraguay, Universidad Nacional de Pilar, Asuncion, Paraguay.

Chavez PA, 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Rem Sens Env 24, 459-479.

Chu YK, Milligan B, Owen RD, Goodin DG, Jonsson CB, 2006. Phylogenetic and geographical relationships of hantavirus strains in eastern and western Paraguay. Am J Trop Med Hyg 75, 1127-1134.

Cohen J, 1960. A coefficient of agreement for nominal scales. Ed Psych Measure 20, 37-46.

Congalton RG, Mead RA, 1983. A quantitative method to test for consistency and correctness in photo interpretation. Photogram Eng Rem Sens 49, 69-74.

Curran PJ, Atkinson PM, Foody GM, Milton EJ, 2000. Linking remote sensing, land cover and disease. Adv Parasitol 47, 37-80.

Emmons LH, Feer F, 1990. Neotropical rain forest mammals: a field guide. University of Chicago Press, Chicago, USA.

Figueiredo MdSL, Fernandez FdS, 2004. Contrasting effect of fire on populations of small rodent species in fragments of Atlantic Forest in Brazil. J Trop Ecol 20, 225-228.

Gentile R, D'Andrea PS, 2000. Population dynamics and reproduction of marsupials and rodents in a Brazilian rural area: a five-year study. Stud Neotrop Fauna Environ 35, 1-9.

Glass GE, Cheek JE, Patz JA, Shields TM, Doyle TJ, Thoroughman DA, Hunt DK, Enscore RE, Gage KL, Irland C, Peters CJ, Bryan RT, 2000. Using remote sensing data to identify areas at risk for hantavirus pulmonary syndrome.

Emerg Infect Dis 6, 238-247.

Glass GE, Shields TM, Parmenter RR, Goade D, Mills JN, Cheek J, Cook J, Yates TL, 2006. Predicted hantavirus risk in 2006 for the Southwestern U.S. Occasional Papers, Museum of Texas Tech University 255, 1-16.

Goodin DG, Koch DE, Owen RD, Chu YK, Hutchinson JMS, Jonsson CB, 2006. Land cover associated with hantavirus presence in Paraguay. Global Ecol Biogeogr 15, 519-527.

Gurney CM, Townshend JRG, 1983. The use of contextual information in the classification of remotely sensed data. Photogram Eng Rem Sens 49, 55-64.

Haralick RM, Shanmugam KS, 1974. Combined spectral and spatial processing of ERTS imagery data.

Hay GJ, Marceau DJ, Dube P, Bouchard A, 2001. A multiscale framework for landscape analysis: object-specific analysis and upscaling. Landscape Ecol 16, 471-490.

Howard SM, Lacasse JM, 2004. An evaluation of gap-filled Landsat SLC-off imagery for wildland fire burn severity mapping. Photogram Eng Rem Sens 70, 877-880.

Huang C, Kim S, Alstatt A, Townshend JRG, Davis P, Song K, Tucker CJ, Rodas O, Yanosky A, Clay R, Musinsky J, 2007. Rapid loss of Paraguay's Atlantic forest and the status of protected areas - a Landsat assessment. Rem Sens Env 106, 460-466.

Iqbal M, 1983. An introduction to solar radiation. Academic Press, Toronto, Canada.

Jansen LLF, van Amsterdam JD, 1991. An object based approach to the classification of remotely sensed images, Proceedings of the International Geoscience and Remote Sensing Symposium (IGARRS) 29, 1171-1176.

Jensen J, 1995. Introductory Digital Image Processing: A Remote Sensing Perspective. Prentice-Hall, Upper Saddle River, NJ, USA.

Kazmi SJH, Usery EL, 2001. Application of remote sensing and GIS for the monitoring of diseases: a unique research agenda for geographers. Rem Sens Rev 20, 45-70.

Kitron U, Clennon JA, Cecere MC, Gürtler RM, King CH, Vazquez-Prokopec G, 2006. Upscale or downscale: applications of fine scale remotely sensed data to Chagas disease in Argentina and schistosomiasis in Kenya. Geospatial Health 1, 49-58.

Linthicum KJ, Bailey CL, Davies FG, Tucker CJ, 1987. Detection of Rift Valley fever viral activity in Kenya by satellite remote sensing imagery. Science 235, 1656-1659.

Lobo A, Chic O, Casterad A, 1996. Classification of Mediterranean crops with multisensor data: per-pixel versus per-object statistics and image segmentation. Int J Rem Sens 17, 49-55.

Martin V, De Simone L, Lubroth J, Ceccato P, Chevalier V, 2007. Prespectives on using remotely-sensed imagery in predictive veterinary epidemiology and global early warning systems. Geospatial Health 2, 3-14.

Mushinzimana E, Munga S, Minakawa N, Li L, Feng CC, Bian L, Kitron U, Schmidt C, Beck L, Zhou G, Githeko AK, Yan G, 2006. Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands. J Milar 5, 13.

Naidoo R, Hill K, 2006. Emergence of indigenous vegetation classifications through the integration of traditional ecological knowledge and remote sensing analysis. Environ Manage 38, 377-387.

Nitikman LZ, Mares MA, 1987. Ecology of small mammals in a gallery forest in central Brazil. Ann Carnegie Mus 56, 75-95.

Ostfeld RS, Glass GE, Keesing F, 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol Evol 20, 328-336.

Pavlovsky, EN 1966. Natural nidality of transmissible disease: with special reference to the landscape epidemiology of zooanthroponoses. University of Illinois Press, Urbana, IL, USA.

Pinzon JE, Wilson JE, Tucker CJ, Arthur R, Jahrling PB, Formenty P, 2004. Trigger events: enviroclimatic coupling of Ebola hemorrhagic fever outbreaks. Am J Trop Med Hyg 71, 664-674.

Richards JA, Jia X, 1999. Remote sensing digital image analysis: an introduction. Springer-Verlag, NY, USA.

Rogers DJ, Myers MF, Tucker CJ, Smith PF, White DJ, 2002. Predicting the distribution of West Nile fever in North America using satellite sensor data. Photogram Eng Rem Sens 58, 809-813.

Sarmiento G, 1983. The savannas of tropical America. In: Bourliere F (Ed). Tropical Savannas Ecosystems of the World, Elsevier, Amsterdam, The Netherlands.

Schmaljohn C, Hjelle B, 1997. Hantaviruses: a global disease problem. Emerg Infect Dis 3, 95-104.

Skole D, Tucker CJ, 1993. Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to

1988. Science 260, 1905-1910.

Song C, Woodcock CE, Seto KC, Lenney MP, Macomber SA, 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Rem Sens Env 75, 230-244.

Williams RJ, Bryan RT, Mills JN, Palma RE, Vera I, De Velasquez F, Baez E, Schmidt WE, Figueroa RE, Peters CJ,

Zaki SR, Khan AS, Ksiazek TG, 1997. An outbreak of hantavirus pulmonary syndrome in western Paraguay. Am J Trop Med Hyg 57, 274-282.

Yu Q, Gong P, Clinton N, Biging G, Kelley M, Shirokauer D, 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. Photogram Eng Rem Sens 72, 799-811.