FOREST COVER ASSESSMENT OF THE MASSIF DE LA HOTTE, HAITI

Zhiqiang Yang & **Erik Haunreiter**, Oregon State University
Warren B. Cohen, USDA Forest Service (Corvallis, Oregon)
Final Report of Subcontract
4 September 2014

Primary CEPF Project:
Ecosystem Threat Assessment and Protected Area Strategy for the Massif de la Hotte Key Biodiversity Area, Haiti
Grant 62132 to Pennsylvania State University
(Stephen B. Hedges, Project Leader; now at Temple University)

CRITICAL ECOSYSTEM PARTNERSHIP FUND

Introduction

This project involves the delineation of original forest cover in the Massif de la Hotte, a region of Haiti especially rich in biodiversity, and is part of a larger CEPF-funded project assessing threats to the ecosystems in this region. Most terrestrial biodiversity is adapted to, and requires, original forest, with its greater plant diversity and higher humidity. Consequently, the distribution of such forest is often used as an indicator of the distribution and health of the native biodiversity. However, it has been estimated that Haiti may have only 1% of its original forests remaining, placing it among the most threatened countries in the world in terms of biodiversity loss.

Forest cover assessments for countries, such as those periodically presented by the Food and Agricultural Organization of the United Nations, normally concern all vegetation types covering the land. Distinguishing the original forest is more difficult, but is necessary to make accurate assessments of ecosystem threats. This is because inclusion of secondary forests, tree plantations, and invasive tree species, will inflate estimates, giving a false impression that the ecosystem is less threatened. In fact, recent reports of forests "increasing" in various countries, and on islands such as Cuba and Puerto Rico, involve the expansion of younger, non-native (invasive) tree species.

For this project, we wanted the most accurate assessment of original forest cover so we used methods, requiring greater efforts, that are not commonly used in reports of "forest cover" in the conservation literature. Here, we define original, native forest (hereafter referred to as original forest) as any area that has not experienced a disturbance and was consistently closed canopy forest during the period of observation.

The information on remotely-sensed forest cover presented here will be used alongside other data gathered in the primary CEPF-funded project to assess ecosystem threats in the Massif de la Hotte Key Biodiversity Area. Those other data include land use activities, ground observations of plant communities and forests, and the distribution of vertebrate species occurring in the region, with emphasis on two major study areas: the Chaine de la Grande Colline (the large mountain range west of Macaya National Park) and Grand Bois mountain near Les Anglais.

Methods

1. Remote Sensing Images

Our study area focused on Massif de la Hotte Key Biodiversity Area (KBA) (Figure 1), which is covered by Landsat WRS2 path 10 row 47 (Figure 2). Since we are using the historical spectral trend to help identify original forest in the study area, we processed a large number of Landsat images from 1984 to 2013 for the study area. We assembled over 200 TM/ETM+/OLI images and converted them to surface reflectance to ready them for analysis.

Figure 1. Study area.

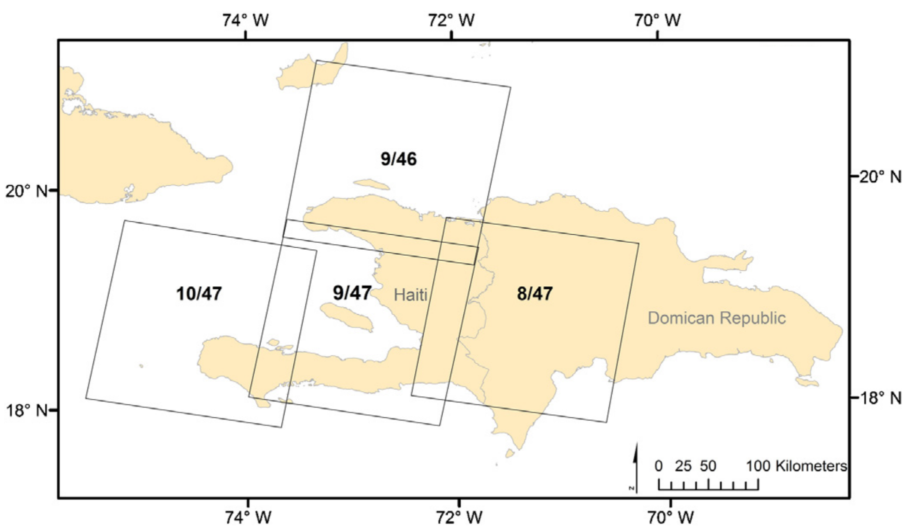

Figure 2. Landsat path/row boundaries for Haiti.

2. Collection of Training Data and Landsat Image Classification

2.1 Training Data

Lacking a suitable sample of current ground information, we relied on aerial photos and high-resolution satellite data available in Google Earth to help identify mature forest cover and develop a training data set for the classification of the Landsat imagery. To facilitate the interpretation, we overlaid a grid with 900m x 900m grid cell size over the study area (Figure 3), and interpreted every fourth cell (every other cell in the row and every other column). Due to cloud cover, and small local sample sizes for forest, as needed we also interpreted pixels outside of the initial selection. The minimum sample unit considered was nine Landsat pixels (approximately 1ha).

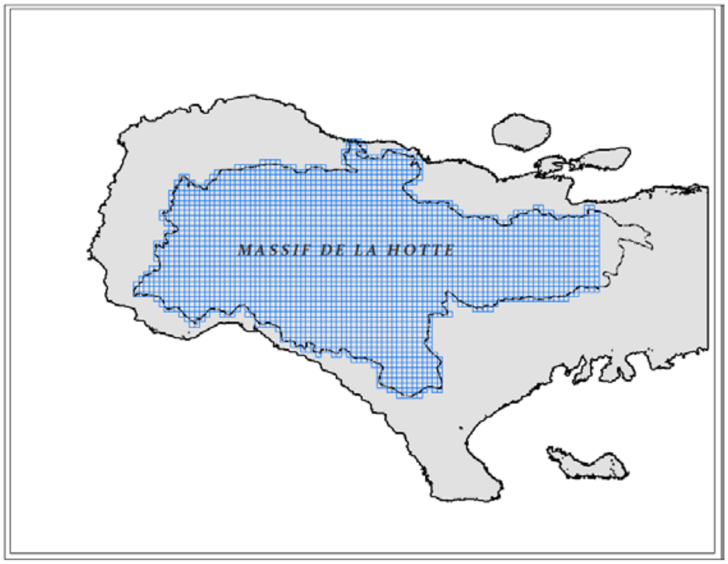

Figure 3. Sampling grid for study area.

To create training data for image classification, we digitized representative polygons of three land use/land cover classes in Google Earth: closed forest, open/degraded forest, and non-forest (Table 1).

Table 1. Training data set.

Class	# of polygons
Closed forest	49
Open forest	88
Non-forest	83

2.2 Cloud Masking and Image Compositing

To map forest cover, our goal was to have one clear view for each 30m pixel every year. Ideally, there would be one fully clear image acquired at roughly the same vegetation phenological state every year across all years. However, the study area is cloudy, with clouds in every image in different locations and some areas being nearly perpetually cloudy. This led to the requirement for selecting only cloud-free pixels to create an image composite for each year of the time series, while minimizing the effect of phenology.

All available images were selected from the archive. To identify clouds in a given image we used the union of the LEDAPS and fMASK cloud masks, as each has a somewhat different set of identified cloudy pixels, enabling use of only the best data for the subsequent composites. The number of images per year ranged from 2 to 21 (Figure 4a), with the percent of cloud free pixels in the composite images throughout the time series being highly variable (Figure 4b).

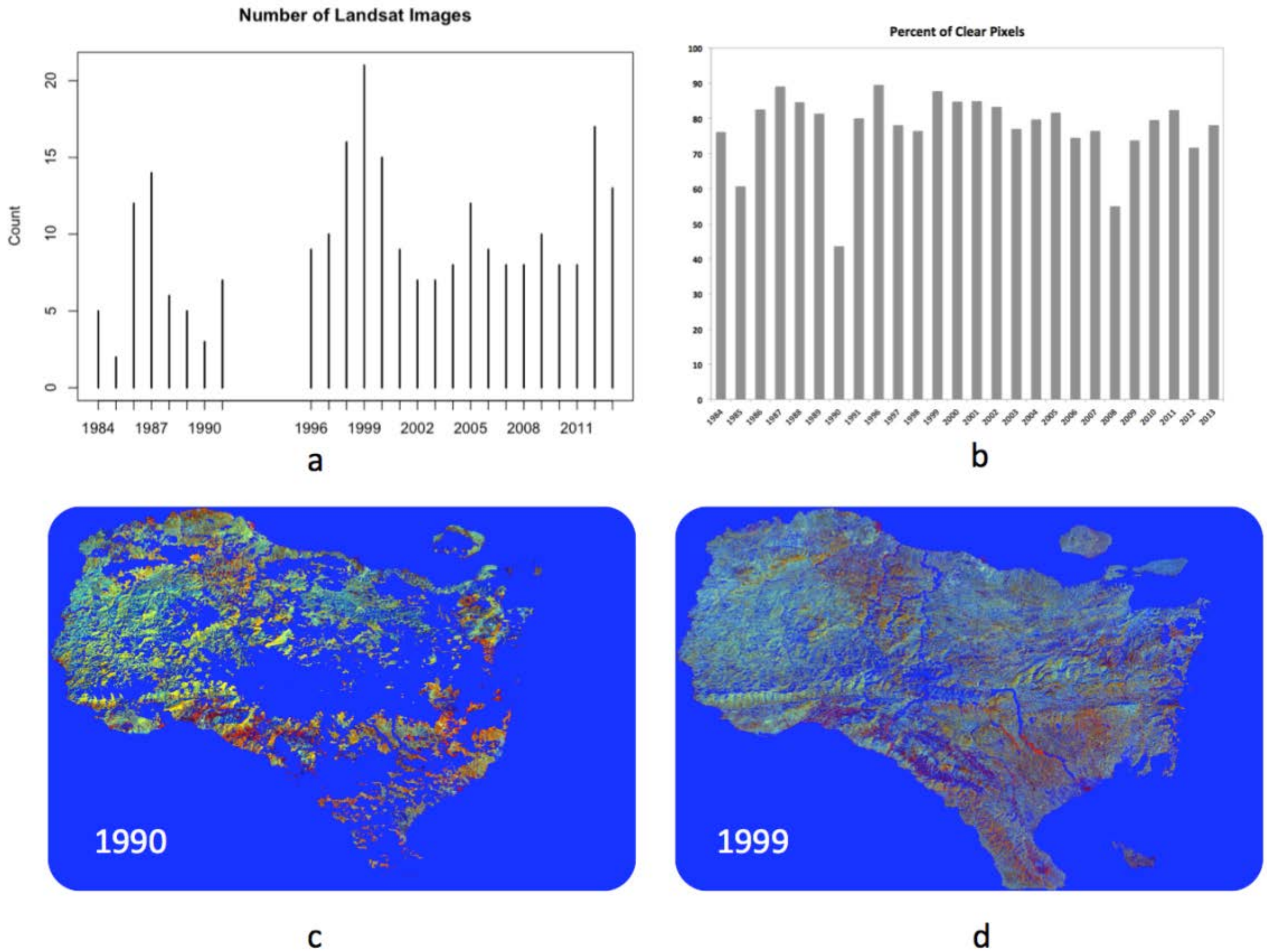

Figure 4. (a) Number of images available for each year; (b) percent of clear pixels in composite image for each year; and images for (c) 1990 and (d) 1999.

To construct the composites for each year, we calculated pixel-wise maximum NDVI values for the given year. This had the effect of both screening residual clouds and minimizing the effects of phenology. Composite spectral values for all the pixels within the training polygons were extracted. Because of residual cloud-masked pixels within each composite image there were portions of polygons and whole polygons that could not be used for training (Table 2).

Table 2. Number of training data polygons and pixels per class after cloud screening and composite creation.

Class	# of polygons	# of pixels
Closed forest	47	1281
Open forest	79	1741
Non-forest	76	1255

2.3 Mapping

The goal of mapping is to identify pixels that are closed forest throughout the time series. The best chance of accomplishing that is dependent upon the ability of the Landsat data to consistently identify closed forest through time. Our approach was based on the classification of each year's composite image and then employing a temporal weighting scheme.

For the classification, we used Random Forest to model the three classes at the pixel level with Landsat composite spectral information. In addition, we incorporated elevation and aspect derived from 30m DEM data (NATHAT mission in June 2010, haitidata.org). The out of bag (OOB) error for the Random Forest model was 7.15% using 500 trees. The confusion matrix derived for the training datasets reveals low prediction error for the closed forest and non-forest classes, with a more moderate level of error for the intermediate class (Table 3).

Table 3. Random forest pixel-level confusion matrix using the training data.

	Prediction			
Reference	**Closed forest**	**Open forest**	**Non-forest**	**Omission**
Closed forest	1198	79	4	0.065
Open forest	58	1610	73	0.075
Non-forest	2	90	1163	0.073
Commission	0.048	0.095	0.062	

Because the Landsat data are calibrated across time (i.e., surface reflectance) the Random Forest model developed with the 2013 training data was applied to the composite data for all years (1984-2013). For each year, two datasets were created: the class map resulting from the assignment to each pixel of the most probable class (left), and the class probability map (right) where probability of each class is shown as an RGB composite (Figure 5).

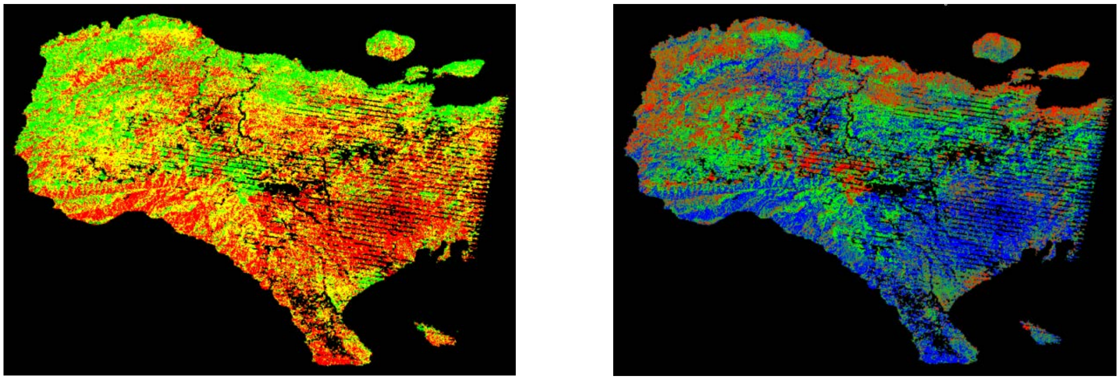

Figure 5. Classification and class probability maps for 2013. (Left) Map showing most probable classes: closed forest (green), open forest (yellow), non-forest (red). (Right) Class probability as an RGB composite, where red is closed forest probability, green is open forest probability, and blue is non-forest probability).

The Random Forest mapping result for any given year (e.g., 2013, Figure 2) only represents the classification for that given year (as the most likely class), without consideration of either the absolute value of the probabilities or historical landscape condition. The 2013 closed forest probability was then combined with the likelihood of being closed forest throughout the time series (1984-2013) to identify pixels having the highest likelihood of being consistently closed forest.

Results

Using the historic closed forest likelihoods as weights assures that pixels labeled as closed forest in 2013 were likely that same class throughout the time period. Using the current probability weight assures that a pixel with recent disturbance in previously closed forest is identified, and thus not mislabeled in 2013 original forest. The resulting map shows the areas where closed, mature forest is most likely to be found (figure 6).

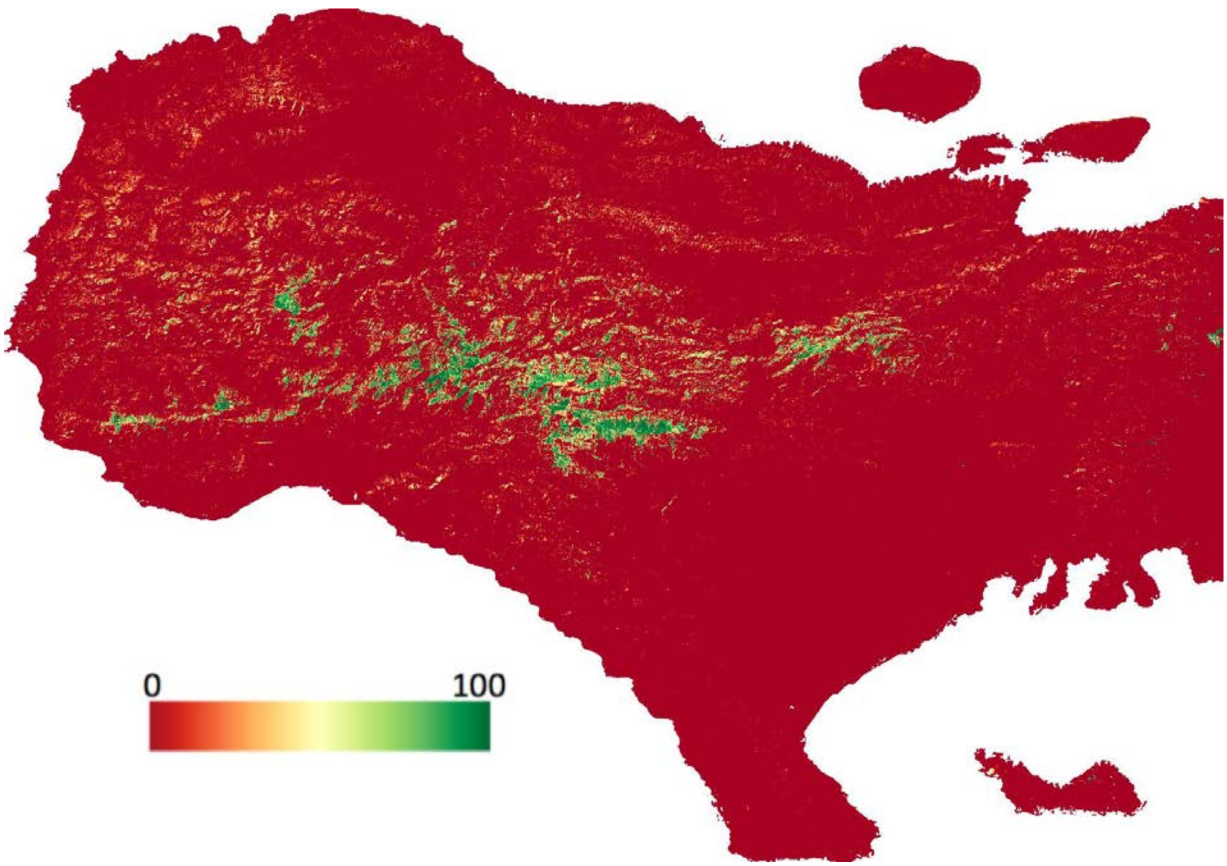

Figure 5. Relative probability of being consistently closed forest. These data are available for download.