Coasian Motion: Forest cover loss and the central tendency to disperse

Dan Hammer, Robin Kraft, David Wheeler

August 20, 2013

Introduction

Tropical deforestation may contribute over 15% of annual greenhouse gas emissions. Any viable effort to mitigate climate change will have to address deforestation. International efforts to curb the rate of deforestation have thus far been composed of a series bilateral agreements, despite calls for a comprehensive, global mechanism. The United Nations has proposed the Reducing Emissions from Deforestation and Degradation (REDD) framework to directly address local incentives to convert forest to fiscally productive lands. This grand Coasian bargain, however, has been frustrated by many factors, including high transaction costs, discord among member countries, and poorly defined property rights. The number of frustrating factors is an increasing function of the number of relevant member countries. The devil is in the details, and the amount of details rises exponentially with the number of contracting agents. As with any network graph, the number of edges increases exponentially with the number of nodes. The UN REDD framework is effectively an attempt to consolidate negotiations; but this effort has met limited success due in part to a constantly shifting natural and political landscape for forest conservation.

This paper demonstrates that the basis for a comprehensive agreement is becoming more complex. Specifically, this paper utilizes satellite imagery to show that the dispersion of forest clearing across administrative units is increasing. The number of countries with a substantial share of forest clearing activities is increasing, potentially forestalling a truly comprehensive conservation framework.

Forest clearing detection

The data on forest clearing activity is based on the classification of stacked images from the Moderate Resolution Image Spectroradiometer (MODIS) sensor aboard NASA's Terra and Aqua satellites. Both time series and cross-sectional characteristics are extracted from the images, and then compared against historical (2000-2005) forest cover loss data (Hansen, 2008). The classification rule is derived from the comparison, and then applied forward to more recent imagery. The final output is a probability of forest cover loss between January 1, 2000 and the acquisition date of the most recent MODIS imagery. The resulting alerts are highly accurate when compared against other, production-level forest clearing alerts. A notable result is that a relatively simple algorithm can produce reasonably accurate alerts, especially to capture broad trends in forest clearing activity.
The algorithm to collect forest clearing alerts consists of four principal steps:

1. Time series characteristics from the derived Normalized Difference Vegetation Index (NDVI) are extracted for each 500-meter pixel in the humid tropics (MOD13A1). These characteristics include long-term trends and variation in the NDVI, along with short-term break detection. The techniques are common in time series econometrics, often in the context of financial indices. The established econometric techniques are repurposed to account for natural cycles.

2. Cross-sectional methods are employed on derived images to identify clusters of the extracted time dependent characteristics. This includes, for example, the largest detected drop in the NDVI series for

a cluster of pixels. The largest, neighborhood drop is ascribed as a pixel-level characteristic, acting as a smoothing operator across the image, much like a moving average over time.

3. The pixel-level attributes from the time series and cross-sectional analysis are compared against historical data on forest cover loss. The precise method of comparison is a logistic, ridge classifier, implemented in parallel to accommodate very large, distributed data.

4. The classification rule is applied to the most recent MODIS imagery for each ecoregion to produce a probability of forest cover loss. The probability can be interpreted as a measure of *strength of signal* of clearing activity. There is substantial noise via measurement error or cloud cover, which may mask the spectral signals associated with forest conversion. The continuous index rises as the signal of forest cover loss persists over time.

The precise algorithm, along with detailed documentation, is stored as an open source repository, with a growing number of contributors. The final output can be fully replicated; every minor step, of which there are many, is open for public review. The result is a highly accurate alerting system of forest cover loss in the humid tropics. There is limited number of publicly available and high quality images for validation. The two primary data sets are (1) the forest cover loss hotspots data that serves as the training data set, and (2) the annual deforestation data in Brazil, released by the national institute for space research (INPE). The data set, called PRODES, reports the boundaries of deforestation for the years 2000 through 2010. This third-party data offers reliable, albeit spatially limited, information on annual land cover change on which to base accuracy assessments, as used by Townshend *et al.* (2004). In line with this simple overlay, the following table compares the PRODES data against both the final output and the pan-tropical training data set. The comparison is limited to the spatial extent of the PRODES data.

Table 1: Accuracy assessment, PRODES versus FCLH and FORMA, 2000-2006 and 2007-2010

	FCLH	FORMA (30)	FORMA (50)	FORMA (80)
2000 to 2006				
Overall accuracy	89.59	89.55	88.89	88.09
Change producer	21.89	22.64	16.28	9.57
Change user	95.33	91.30	95.41	97.91
No change producer	99.84	99.67	99.88	99.97
No change user	89.41	89.49	88.74	87.96
2007 to 2010				
Overall accuracy	-	85.78	85.69	85.35
Change producer	-	13.53	10.84	7.65
Change user	-	78.72	87.34	93.38
No change producer	-	99.32	99.71	99.90
No change user	-	85.98	85.66	85.24

[Analysis of table, followed by analysis of hansen for full tropics. first show out of sample validity, and then, after, show the less convincing in-sample validity for a greater spatial extent (outside of Brazil).]

Dispersion

The alerts are aggregated into three levels of administrative units: country, province, and sub-province. These units correspond to levels 0, 1, and 2, respectively, in the Global Administrative boundaries (GADM) data. The aggregated alerts can be aggregated by a simple count, a sum of probabilities, or a combination of the two. Each method offers a slightly different view onto the data. Here, we offer only the simple count, but the full set of results is offered on the companion code repository, which serves as a robustness check for the reported results. Specifically, an alert is identified when the probability of forest cover loss exceeds a pre-defined confidence threshold. A natural threshold is 50%, as it represents the steepest point on the logistic function, which underlies the classification rule. The 50% threshold is easily defensible and defines the alerts, which are aggregated for each administrative unit and for each 16-day interval.

The pantropical rate of industrial-scale forest clearing alerts are reported in Figure 1. The seasonality of tropical alerts is clear; and through the seasonality, so too is the upward trend of forest clearing activity. Equally pertinent to a comprehensive conservation framework, however, is the spatial distribution of the clearing activity. A common measure of dispersion is the normalized Shannon Entropy measure (citation needed). A value of 1 indicates pure entropy or dispersion, associated with a uniform distribution across units. Alternatively, a value of 0 indicates that the frequency distribution consists of a single value: no dispersion at all.

Let i be the pixel index, and j be the set of pixel indices in administrative unit j. The aggregate level of clearing activity in administrative unit j and time period t is given by Equation (**??**).

$$D_{jt} = \sum_{i \in C_j} \mathbb{I}(p_{it} \geq 0.5) \cdot p_{it} \tag{1}$$

We characterize the spread of clearing activity across all administrative units in the sample by using Shannon's entropy criterion, defined in equation Y.

$$E_t = -\sum_{j=1}^{n} \frac{D_{jt}}{D_t} \log_2 \frac{D_{jt}}{D_t} \quad \text{with} \quad D_t = \sum_j D_{jt} \tag{2}$$

E ranges from 0 to 1, and larger values of E indicate higher dispersion across units. A value of 1 is a uniform distribution, with equal values for all units.

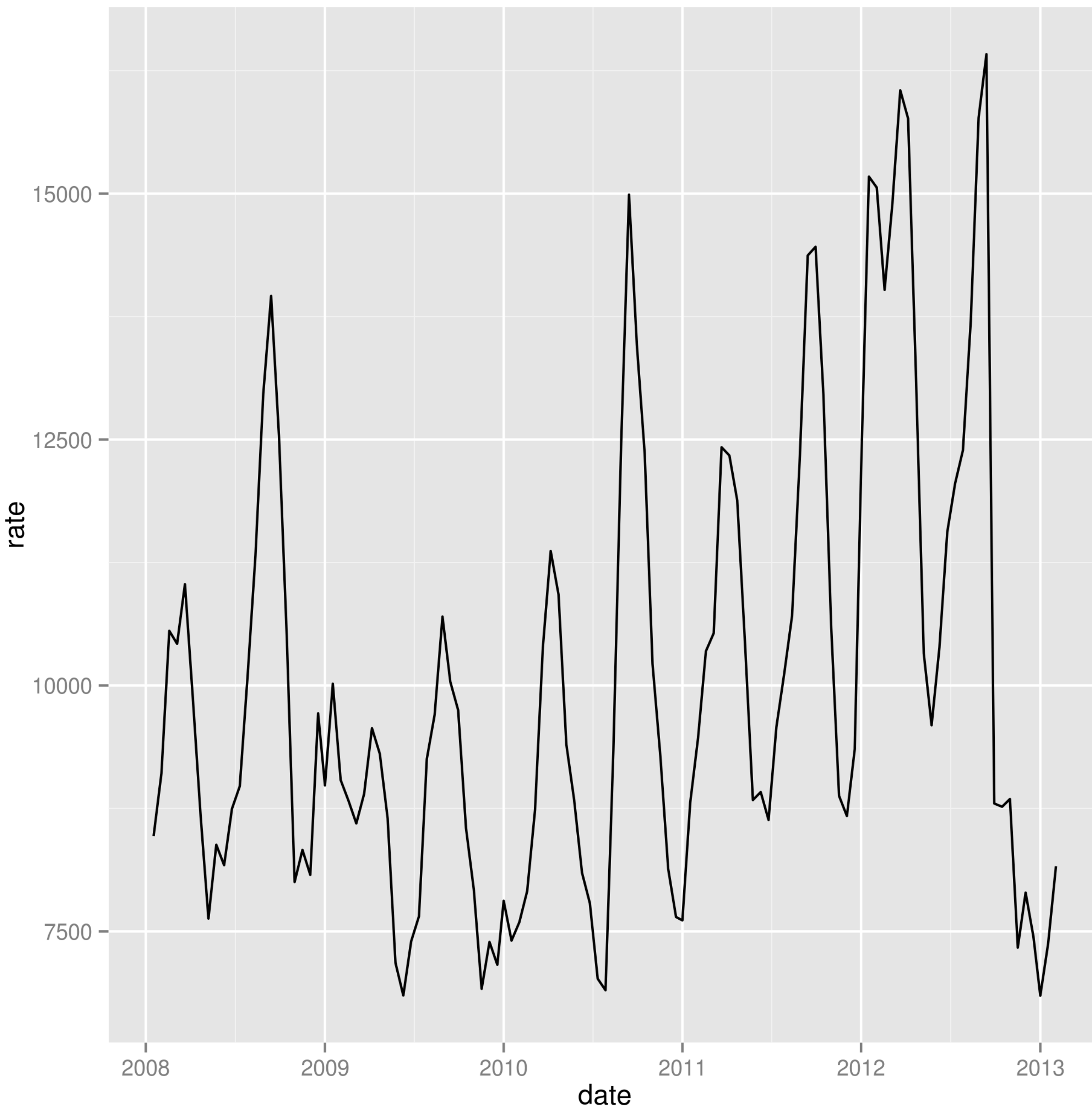

Figure 1: Pan-tropical Rate

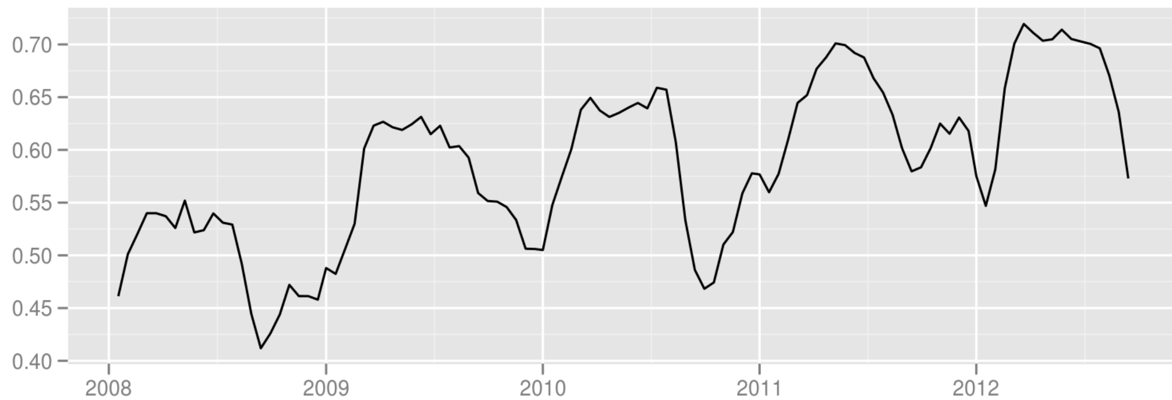

Figure 2: Country level dispersion

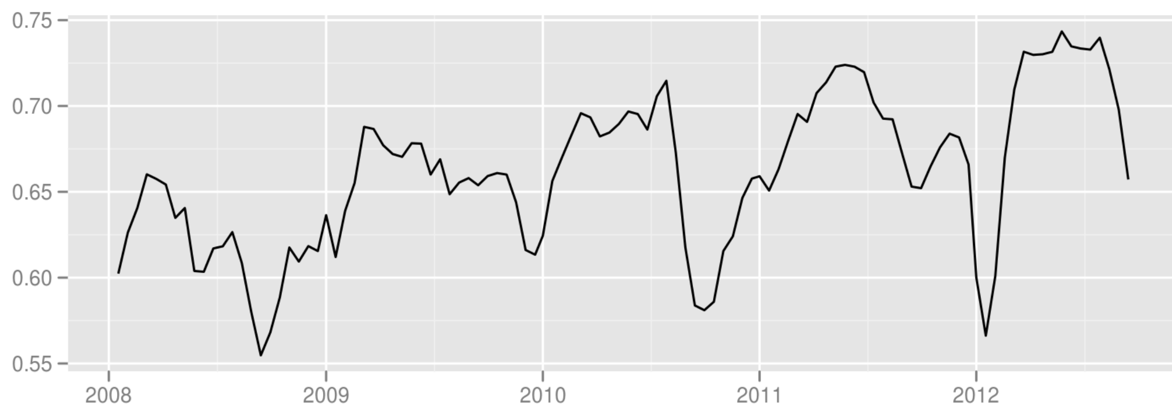

Figure 3: Province level dispersion