FORMA: Forest Monitoring for Action—Rapid Identification of Pan-tropical Deforestation Using Moderate-Resolution Remotely Sensed Data

Dan Hammer, Robin Kraft, and David Wheeler

Abstract

Rising concern about carbon emissions from deforestation has led donors to finance UN-REDD (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries), a program that offers direct compensation for forest conservation. Sustainable operation of UN-REDD and other direct-compensation programs will require a transparent, credible, frequently updated system for monitoring deforestation. In this paper, we introduce FORMA (Forest Monitoring for Action), a prototype system based on remotely sensed data. We test its accuracy against the best available information on deforestation in Brazil and Indonesia. Our results indicate that publicly available remotely sensed data can support accurate quarterly identification of new deforestation at 1 km spatial resolution. More rapid updates at higher spatial resolution may also be possible. At current resolution, with efficient coding in publicly available software, FORMA should produce global updates on one desktop computer in a few hours. Maps of probable deforestation at 1 km resolution will be accessible with Google Earth and Google Maps, with an open facility for ground-truthing each pixel via photographs and text comments.

FORMA:

Forest Monitoring for Action—Rapid Indentification of Pan-tropical Deforestation Using Moderate-Resolution Remotely Sensed Data

Dan Hammer
Robin Kraft
David Wheeler

Center for Global Development
1800 Massachusetts Ave., NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

FORMA: FOREST MONITORING FOR ACTION

RAPID IDENTIFICATION OF PAN-TROPICAL DEFORESTATION
USING MODERATE-RESOLUTION REMOTELY SENSED DATA*

DAN HAMMER
ROBIN KRAFT
DAVID WHEELER

ABSTRACT. Rising concern about carbon emissions from deforestation has led donors to finance UN-REDD (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries), a program that offers direct compensation for forest conservation. Sustainable operation of UN-REDD and other direct-compensation programs will require a transparent, credible, frequently updated system for monitoring deforestation. In this paper, we introduce FORMA (Forest Monitoring for Action), a prototype system based on remotely sensed data. We test its accuracy against the best available information on deforestation in Brazil and Indonesia. Our results indicate that publicly available remotely sensed data can support accurate quarterly identification of new deforestation at 1 km spatial resolution. More rapid updates at higher spatial resolution may also be possible. At current resolution, with efficient coding in publicly available software, FORMA should produce global updates on one desktop computer in a few hours. Maps of probable deforestation at 1 km resolution will be accessible with Google Earth and Google Maps, with an open facility for ground-truthing each pixel via photographs and text comments.

1. INTRODUCTION

Forest clearing is an enormous contributor to global warming, accounting for some 15% of annual greenhouse gas emissions [17]. Most forest clearing occurs in developing countries that have limited resources and regulatory capacity. Since these countries understandably focus their energy and resources on poverty alleviation, their support for forest conservation will be weak as long as forested land has a higher market value in other uses. Under these conditions, many actors will continue clearing their forested land unless they are given conservation payments that match or exceed the opportunity cost of the land. This economic insight has led the UN to establish UN-REDD (Reducing Emissions from Deforestation and Forest Degradation in Developing Countries), a program that helps countries prepare for an eventual direct compensation scheme for forest conservation. The first prototype for REDD operations is the World Bank's Forest Carbon Partnership Facility (FCPF), launched at the UN's Bali conference on climate change in December, 2007. Target capitalization for this prototype facility is over $300 million [16]. However, the UNFCCC estimates that full conservation of remaining forests in the tropics and subtropics will require $12.2 billion annually [15].[1] A compact negotiated this year in Copenhagen may support an expansion of UN-REDD to this scale,

[1]UNFCCC (2007) defines the needed financial flow as the opportunity cost of forested land in the most profitable alternative use. Alternative uses, or deforestation drivers, include cattle ranching, small-scale agriculture, shifting cultivation, and gathering fuelwood and non-timber forest products. The analysis assumes that without conservation payments, deforestation/degradation will continue at 12.9 million hectares/year, emitting 5.8 Gt of CO_2. It maintains the current hectare proportions for each driver,

because carbon emissions abatement from forest conservation is much lower-cost than abating emissions from fossil fuels (Stern, 2006). The UNFCCC's estimate of CO_2 emissions from forest clearing (5.8 Gt) implies an average abatement cost of only \$2.10/tonne (at an annual payment of \$12.2 billion).

Sustained international support for such enormous payment flows – equal to about 10% of existing development aid – will hinge on the operational credibility of REDD programs. For accountability, the global community will need access to a monitoring system that provides detailed, accurate and timely identification of deforestation in conservation-payment areas. To ensure the broadest access and credibility, the monitoring system should be truly transparent and reproducible, making data, algorithms and processing workflows publically available and usable in free or inexpensive software. Its outputs should be automatically converted into detailed, easy-to-understand displays accessible with a web browser or similar free software, and would ideally include a public facility for both casual and systematic ground-truthing through geolocated photographs and commentaries.

In this paper, we describe the construction and testing of a prototype system that meets these conditions. Called FORMA (Forest Monitoring for Action), the system utilizes moderate-resolution data recorded daily by the Moderate Resolution Imaging Spectrometer (MODIS), which operates on NASA's Terra and Aqua (EOS PM) satellite platforms. MODIS data products going back as far as February 2000 are freely available at varying resolutions. Although its signal-processing algorithms are relatively complex, FORMA is based on a common-sense observation: Tropical deforestation involves the burning of biomass and a pronounced temporary or long-term change in vegetation color, as the original forest is cleared and replaced by pastures, croplands or plantations.

Accordingly, FORMA constructs deforestation indicators from MODIS-derived data on the incidence of fires and changes in vegetation color as identified by the Normalized Difference Vegetation Index (NDVI).[2] It then calibrates to local deforestation by fitting a statistical model that relates the MODIS-based indicator values to the best available information on actual deforestation in each area. FORMA incorporates biological, economic and social diversity by dividing the monitored territory into 100 km^2 blocks and separately fitting the model to data for the 10,000 1 km^2 parcels in each block.[3] The dependent variable for each pixel is coded "1" if it has actually experienced deforestation within the relevant time period, and "0" otherwise. The MODIS-based indicator values are the independent variables. For all tropical countries except Brazil, the best identification of recent deforestation has been published in *Proceedings of the National Academy of Sciences* by Hansen, et al. (2008), who estimate the incidence of deforestation for 500m parcels in the humid tropics. We calibrate FORMA using the map of forest cover loss hotspots (henceforth referred to as the FCLH dataset) published by Hansen, et al. for the period 2000-2005.[4] In Brazil, higher resolution estimates are also available annually from the INPE PRODES program (2009). We use these estimates to test the accuracy of our FCLH-based calibration for Brazil.

Using the FCLH pan-tropical dataset for 2000-2005, FORMA fits the calibration model to 10,000 observations on deforestation in each 100km^2 block of humid tropical forest area. It then applies the fitted model to monthly MODIS indicator data for the post-2005 period, Q1 2006 to Q4 2008. The output for each month is a predicted deforestation probability for each 1 km^2 parcel outside of previously-deforested areas, as identified in the FCLH map. Monthly observations include significant "noise" introduced by random technical problems, cloud cover, etc. To provide a clearer signal, we smooth the monthly probabilities to provide estimates of likely deforestation on a quarterly basis. The final output is color-coded by probability level

---

and applies the relevant opportunity cost to each part. The result is an estimated total annual compensation payment of \$12.2 billion.

[2]A future version of FORMA will switch from the NDVI to a more recent product of the MODIS Science Team, the Enhanced Vegetation Index (EVI). At the outset, we chose NDVI because we anticipated the need for long time series that would join NDVI to data from MODIS' predecessor, the AVHRR (Advanced Very High Resolution Radiometer). This no longer seems necessary, so a future switch to EVI seems warranted.

[3]See Section 3.3 for a planned improvement in sample definition based on ecoregions as defined by scientists at the World Wildlife Fund (WWF). A detailed description of the terrestrial ecoregions is available at http://www.worldwildlife.org/science/ecoregions/item1267.html

[4]It is important to note that the FCLH data are estimated, not directly observed.

and mapped to identify areas where new fires and rapid vegetation color changes indicate a high probability that deforestation is occurring.

Although we believe that our FORMA prototype has advanced the state of the art in significant ways, we should note at the outset that it represents a proof of concept rather than a finished, optimized system. Our modeling and programming algorithms could undoubtedly be improved further by expert specialists, and we hope that such improvements will be forthcoming once we publicly release our first-generation model.

The remainder of the paper is organized as follows. In Section 2, we discuss previous applications of remote sensing to deforestation; identify shortfalls from the perspective of global monitoring; and relate these to FORMA's design principles. Section 3 introduces the data that we use to calibrate FORMA in each territorial block; develops our statistical probability model; and provides illustrative results for 100 km^2 blocks in Pará, Brazil and Sumatra, Indonesia. We estimate model parameters for the 1 km^2 parcels that experienced deforestation during 2000-2005 according to the FCLH map developed by Hansen, et al. Section 4 tests FORMA's accuracy for Brazil by projecting annual deforestation for 2006-2008 from the MODIS indicators, and comparing the results with deforestation identified by the Brazilian government's PRODES system. In Section 5, we provide an illustrative quarterly application for deforestation since 2005 in Indonesia's Riau Province (central Sumatra). Section 6 introduces the next steps in our research: expansion of FORMA to global coverage of humid tropical forests, starting with Brazil and Indonesia; recalibration of FORMA with improved datasets as they become available, globally or locally; integration of FORMA deforestation probability maps with Google Earth and Google Maps, with an open facility for ground-truthing in each 1 km^2 parcel; estimation of resource requirements for a full global implementation of FORMA; and a tentative prospectus for an independent global facility that will maintain and operate FORMA. Section 7 provides a summary and conclusions.

2. Identification of Deforestation with Remotely Sensed Data

Estimates of deforestation based on remotely sensed data have been available in various forms for decades. Perhaps the most impressive contribution has been made by Brazil's PRODES (2009), which has provided yearly maps of Amazonian deforestation since 1988. Since 2004, these have been augmented by twice-monthly estimates from Brazil's DETER system.[5] Another noteworthy Brazilian contribution is Imazon's Forest Transparency Initiative, which has utilized MODIS data to produce and rapidly disseminate information about deforestation in Mato Grosso State.

Several global-scale deforestation studies have been reported in scientific journals. Although they have laid the groundwork for global monitoring, these studies have not replicated the Brazilian contribution by providing updated, online reporting. Nor are they accessible to non-specialists who do not have a deep understanding of Geographic Information Systems (GIS) and remote sensing techniques. As Grainger (2008) has noted, tracking the long-term trend in tropical deforestation has been problematic. Hansen, et al. (2008) identify global deforestation in humid tropical forests using MODIS and Landsat images for the period 2000-2005. Mulligan (2008) uses remotely sensed data for assessment of land use changes in and around protected areas from 2000 to 2005. Carroll, et al. (2006) identify changes in vegetation cover from 2001 to 2005.

Several institutions provide detailed information on deforestation with varying quality, but they have not attempted continuous global monitoring at high resolution. The FAO (2005) provides a detailed Global Forest Resources Assessment at the country level, updated at 5-year intervals. The World Resources Institute has published detailed maps of deforestation hotspots in Latin America, Asia and Africa for the period 2000-2006. The website maintained by Global Forest Watch[6] provides global information, but with non-standardized spatial and temporal coverage of different datasets by country, infrequent updates, and a map interface that does not permit integrated global views.

In summary, outside of Brazil, currently-available information is not sufficient to ensure transparency, credibility and sustainability for programs like UN-REDD. As Mayaux, et al. (2008) have noted:

---

[5]Detailed descriptions of PRODES and DETER are available from Brazil's National Institute for Space Research (INPE) at http://www.obt.inpe.br/prodes and http://www.obt.inpe.br/deter.

[6]Available at: http://www.globalforestwatch.org/english/index.htm.

National, international and academic institutions have had difficulties delivering accurate information in a way that is useful and relevant for policy formulation, implementation and follow-up. One prominent reason has been the inability of the agencies concerned to establish a commonly accepted, independent, cost-effective and long-term mechanism to deliver remote sensing data to users. Such a framework is still needed.

In addition to Mayaux's criteria, we believe that an optimal system should be global in coverage at "reasonable" spatial resolution, updated frequently, easy to use, low-cost, and easily understood by non-experts. We have designed FORMA to meet these conditions. The raw data files that feed into FORMA are freely downloadable from the sites indicated in Table (1). Processing the MODIS FIRMS data on the incidence of fires is inherently fast, because the underlying observations are relatively sparse in each time period. Changes in vegetation color at very high spatial resolution are another matter, because the supporting database is potentially huge. We have chosen to use the MODIS 1 km NDVI (Normalized Difference Vegetation Index) dataset (MOD13A3), which provides indicators that can be updated monthly at a marginal cost approaching zero, with spatial resolution appropriate for open-access monitoring of UN-REDD-type programs. Higher spatial resolution is of course attainable, but it would be more expensive computationally.[7] We are satisfied with 1 km resolution for this initial version of FORMA, as the raw input datasets can be processed and stored on one desktop computer using industry-standard GIS and statistical software packages.

Table 1. FORMA Database

Data Element	Sensor	Satellite	Spatial	Temporal
NDVI	MODIS	Terra	1 km	Monthly
FIRMS active fire data	MODIS	Terra & Aqua	1 km	Daily
Precipitation Reconstruction over Land	n/a	n/a	0.5 deg	Monthly
Vegetation Continuous Fields	MODIS	Terra	500 m	One-off
GlobCover	MERIS	Envisat	300 m	One-off
FCLH	MODIS	Terra	500 m	One-off

3. The FORMA Model

FORMA processes information in four stages. Stage 1 generates information on short- and long-term trends in vegetation color from MODIS NDVI data for each pixel. We improve the color trend estimates by controlling for variations in local precipitation – as recorded in the PRECL (Precipitation Reconstruction over Land) database – and local seasonal color variations – proxied by monthly dummy variables. Stage 2 combines the NDVI trend information for 2000-2005 with information on fires from the MODIS FIRMS database, for the same pixel during the same period. Then, for each pixel, it matches the NDVI and FIRMS variables to data on deforestation for 2000-2005 [7]. Stage 3 estimates the probability model for territorial blocks that contain many pixels (10,000 per block in the current specification). The dependent variable is coded "1" (FCLH deforestation from Hansen, et al. (2008) during 2000-2005) or "0" (no FCLH deforestation during that period). To ensure the best fit, we drop data for pixels with less than 25% tree cover in 2000 as measured by the Vegetation Continuous Field dataset used by Hansen, et al. (2008). Moving to stage 4, we use the European Space Agency's MERIS (Medium Resolution Imaging Spectrometer) GlobCover dataset to limit estimation to areas identified as forested in 2005. We then use the estimated model parameters and post-2005 monthly data from NDVI (vegetation color) and FIRMS (the incidence of fires) to generate quarterly deforestation probabilities since 2005 for each pixel.

The assumptions underlying our approach are as follows for each pixel:

---

[7]We have designed FORMA for low-cost, timely global detection of deforestation hotspots, and we recognize that detecting some forms of local degradation in small parcels or project areas may require higher-resolution data from Landsat and other systems.
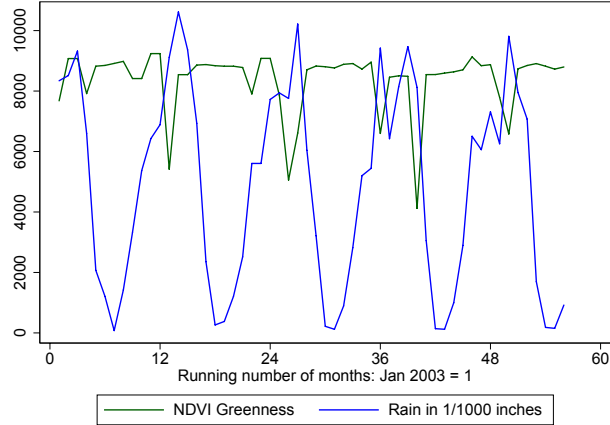
FIGURE 1. NDVI and rain for a pixel in Pará, Brazil (Jan 2003 - Aug 2008)

   i. Forest cover loss (FCL) associated with significant deforestation causes an extra-seasonal decrease in the NDVI vegetation index that is detectable at 1 km resolution. FCL of this magnitude cannot occur without such a decrease, although any specific drop could have another explanation (i.e. deforestation-related decreases in NDVI are a subset of all NDVI decreases).

  ii. More intense deforestation is associated with a higher incidence of fires, as biomass is cleared for land conversion.

 iii. Deforestation is more likely when NDVI decreases and fires are occurring in neighboring pixels, since significant deforestation generally occurs over areas larger than 1 km^2.

3.1. **NDVI Trend Analysis.** The MODIS time series data for each pixel contribute several indicators to estimation of the probability model (Stage 4). From the NDVI, we have extracted both long- and short-run trends. In addition, we are continuing to experiment with measures of shifts in natural cycles around the long-term trend.

3.1.1. *Long-Term NDVI Trend.* The long-term trend is estimated from a regression on the full time series.[8] We analyze areas that were forested in 2000, so a significant decline in the NDVI is a powerful indicator of deforestation. For each pixel i, we extract the linear trend from the following regression equation, fitted to monthly data for 2000-2005:

$$(1) \qquad \text{NDVI}_{it} = \alpha_{0i} + \alpha_{1i}t + \alpha_{2i}rain_{it} + \sum_{m=1}^{11} \eta_{mi}\delta_m + \epsilon_{it},$$

where $\epsilon_{it} \sim N(0,\sigma_i)$; t is time elapsed (in months) since the first month; $rain_t$ is rainfall in month t; and δ_m is a dummy variable for month m.[9] The parameter of interest is α_1, the long-term trend of the NDVI. For probability model estimation, we retain both the magnitude of α_1 and the measure of significance provided by its associated t-statistic. To illustrate, Figure (1) presents monthly NDVI values and rainfall from January 2003 to August 2007, for a pixel in Pará, Brazil.[10] Both rainfall and vegetation color are seasonal, so a substantial portion of the variation in rainfall and vegetation color is captured by the monthly dummies. However, there is clearly extra-seasonal variation in rain, which will affect vegetation color. In

---

[8]As more data become available, we may consider a standard estimation interval of about six years for global estimation, or an interval that is optimized for the conditions in each ecoregion and country.

[9]The effect for the twelfth month is absorbed by the regression constant (α_0).

[10]The graph suggests that lagged rain provides a better fit to fluctuations in NDVI than contemporaneous rain. We are investigating a re-specification of the time-trend model to incorporate different lag structures and precipitation trends.

addition, controlling for the effect of rainfall also incorporates possible long-run climate trends that will affect rainfall – and therefore the vegetation index – over time.[11]

Although the linear NDVI trend is simply derived, our probability model estimation (Section 3.3) indicates that it is a very powerful indicator of deforestation, especially in Brazil. Figure (2a) illustrates this with a map of deforestation in Pará, Brazil from 2000-2008 as identified by the Brazilian space agency's PRODES program. PRODES-identified deforestation is outlined in black, and each pixel is color-coded in shades of brown, with darker shades assigned to more negative t-statistics for the linear NDVI trend. Additional variables add power to the probability model, as we will show in the following sections, but the explanatory power of the NDVI long-term trend is clear from the map of statistical significance – nearly all the PRODES deforestation areas are colored deep brown. Nonetheless there is substantial "noise" in the variation of NDVI values, particularly in Indonesia [Figure (2b)], where there are many isolated pixels with a statistically significant decline in the NDVI but without deforestation. In Figures (2c) and (2d), we map the standard deviation of the NDVI for the Pará subsample over the entire study period (2000-2008) and for the year 2005, respectively, showing that substantial, misleading variation occurs outside of areas identified as deforested during the study period.

3.1.2. *Short-Term NDVI Trend.* The second component of our time series analysis comes from a routine that searches for discrete, semi-persistent drops in the NDVI. In Brazil, long-term trend analysis might suffice because the pattern of land-use change typically involves conversion of forests to cattle pastures or sugar or soy plantations with significantly lower NDVI values and different seasonal variation. In Indonesia, however, the situation is notably different: Forests are often converted to oil palm or acacia plantations, whose NDVI values are close to the original forest values. In oil palm plantation conversions, the NDVI falls initially but quickly returns to near-original levels within one to two years. Figure (3) illustrates the clear difference between typical Brazilian and Indonesian cases. Figure (3a) presents the 32-day composite NDVI values for a deforested Brazilian pixel, while Figure (3b) presents the NDVI values for an Indonesian pixel that has been converted to an oil palm plantation.

Simple estimation of a long-term trend can easily miss the large, semi-persistent drop illustrated in Figure (3b). To capture its effect, we construct a routine that searches for such drops in the NDVI series. First we convert the seasonally-adjusted NDVI series for each pixel into successive time series blocks of length l, so that each series is indexed by $b \in \{1, 2, \ldots, (M - l)\}$, where M is the total number of months. For February 2000 to December 2008, inclusive, there are 107 months ($M = 107$). The length l is chosen to be greater than 12 months so that it will include (and neutralize the effect of) annual cycles in vegetation color. We then estimate a variant of Equation (1) for each pixel and for each time series block, and collect the associated time-trend value α_{1ib}. We compute a multi-year moving average of the time-trend values, with the number of time-trend values given by n. Note that both n and l are effectively temporal smoothing parameters, which we fix for all pixels. We extract the minimum (most negative) moving average value of the time trends, and include the value as another NDVI indicator in the probability model. For the sake of explicit exposition, we assign $maxdrop_i$ to each pixel i, defined by:

$$(2) \qquad maxdrop_i = \min\left\{ \left[\frac{1}{n} \sum_{j=0}^{n-1} \left(\alpha_{1i(b+j)} \right) \right]_b \right\},$$

where we fix $n = 5$ and $l = 15$ for this first version of FORMA. Inclusion of $maxdrop$ in the probability model makes a noticeable difference for the quality of fit, particularly for Indonesia: far more of the variation in deforestation is picked up by the FORMA filter.

We have also experimented with algorithms that seek to capture other components of the NDVI time series for individual pixels. One measure stems from our observation that the variance in the NDVI series

---

[11]Because it relies on NDVI and fires data, FORMA may not clearly distinguish between human-induced deforestation and severe forest stress during periods of prolonged drought and forest "browning", accompanied by naturally-occurring fires. The system will likely generate more "false positive" deforestation signals during such periods, though it is not clear what drought "looks like" in a given region as compared to forest clearing. This is one reason why FORMA's web presentation system will include a facility for pixel-specific ground-truthing by local stakeholders.
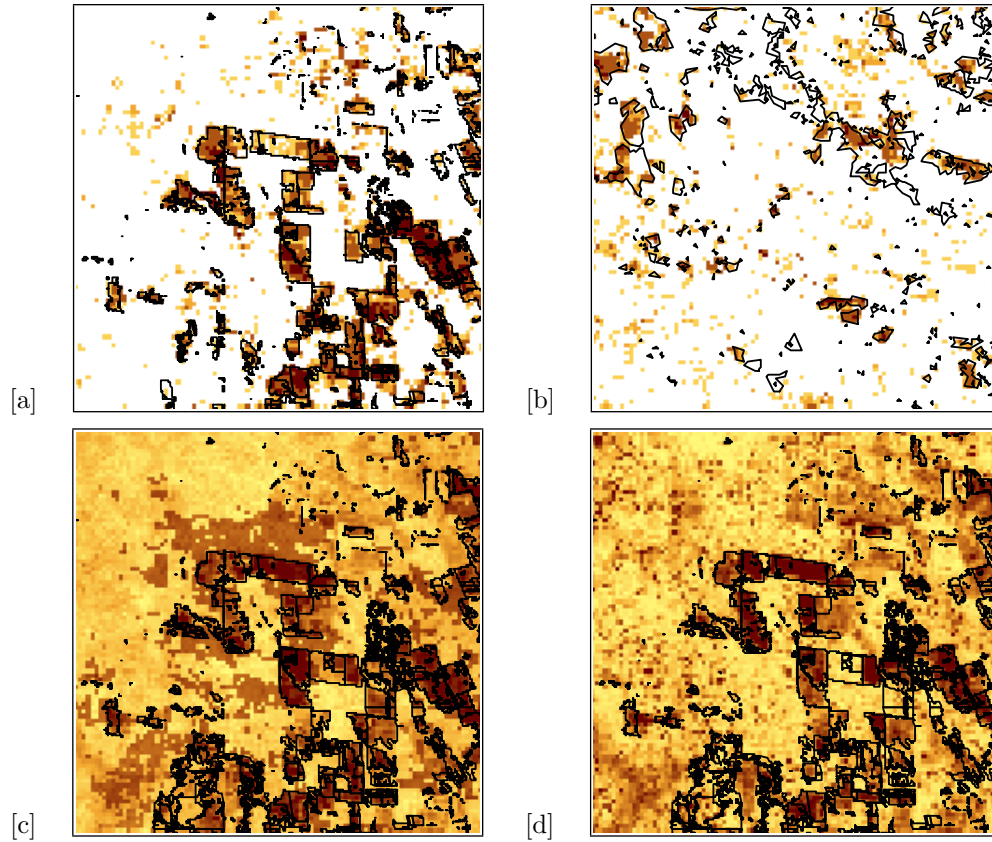
FIGURE 2. NDVI variation, PRODES, fires and t-statistics for Riau and Pará subsamples (Jan 2000 - Dec 2005): [a] PRODES (black outlines) and t-statistics colored from light to dark brown in increments of statistical significance for Pará [b] FCLH (black outlines) and t-statistics for Riau [c] Standard deviation of monthly NDVI, 2000-2008 for Pará, with PRODES (black outlines) [d] Standard deviation of monthly NDVI, Jan - Dec 2005, for Pará, with PRODES (black outlines).

is noticeably altered after a change in land-use, fluctuating wildly compared to its historical behavior. To test for significance, we have calculated ratios of maximum and minimum values of coefficients of variation for the rolling time series blocks used in the previous computation. This has not proved to be significant in probability model estimation, so it has been excluded in the current specification. We are also investigating other possibly-significant dimensions of variation in the NDVI series. Current lines of inquiry include neural network analysis and the use of macroeconomic business cycle filters for more sophisticated identification of natural cycles in the data – and departures from these cycles that may be significantly related to deforestation.[12]

Equations (1) and (2) represent approximately 100 time series regressions for each pixel, so the magnitude of the required data processing appears intimidating at first glance. Global application of FORMA will cover nearly 20 million 1 km^2 parcels, necessitating some 2 billion linear regressions. This is a non-trivial proposition for a standard desktop computer running Stata. But in keeping with our objective of accessibility

---

[12]Additional experimentation could attempt to further generalize the model, capturing all relevant signals from the selected satellite data. Neural network analysis could potentially add explanatory power by generalizing the functional forms; business cycle filters and other macroeconometric techniques may further distill meaningful variation for short-term NDVI analysis.
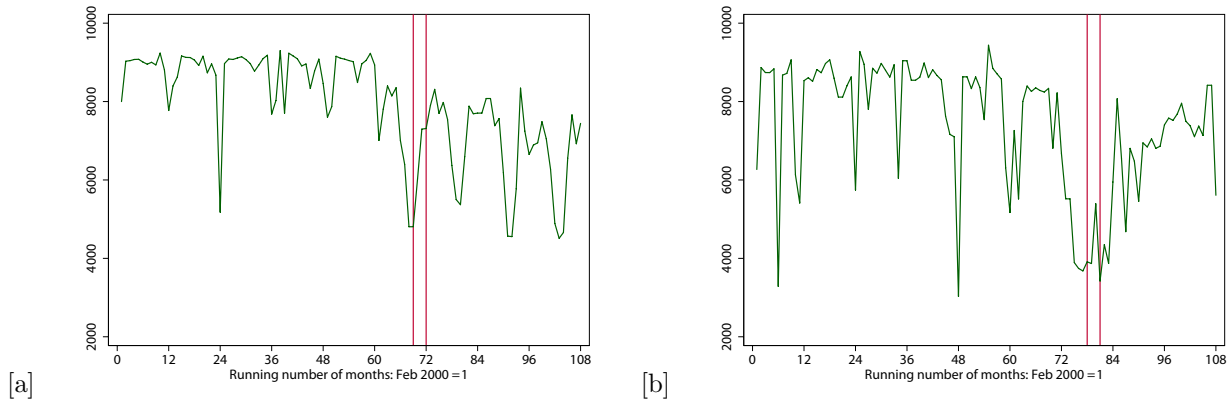
FIGURE 3. NDVI time series for a selected pixel in [a] Pará, and [b] Riau (Feb 2000 - Dec 2008); vertical lines indicate the quarter when deforestation registered in FORMA.

and low cost, we have constructed a system that re-runs the FORMA analysis in a matter of hours on a single computer.[13]

3.2. **Fires Analysis.** The fires time series may also contain multiple components that can be extracted. At present, our probability model only includes the count of fires within a pixel over the relevant time period. We include all recorded illumination events, regardless of brightness or confidence levels. In the next round, we will incorporate the findings of Morton, et al. (2008) by investigating separate counts for fires with greater than 330K brightness, and repeated fires within a 12-month period. The number of high-frequency fires, as opposed to a simple count, may provide a superior signal for clearance of live forests or felled trees, either at the beginning or end of a typical deforesting season. The current measure of fires does not incorporate the timing of fires, which has been shown to carry significant weight in identifying deforestation [11].[14] In the same vein, a large proportion of FORMA's false-positive predictions are associated with isolated, weak illumination events included as fires in our undifferentiated count. Many of these events are undoubtedly spurious, and adoption of a more sophisticated measure should remove many of them in the next iteration of the model.

Figure (4) presents fires overlaid with deforestation for the Riau, Indonesia and Pará, Brazil subsamples. The spatial fit is apparent, especially in Pará, where most of the fires occur within the bounds of deforestation. The spatial fit is less robust for the Riau subsample; there is far less covariation between fires and FCLH deforestation estimated by Hansen, et al. (2008). The probability model, described in the next section, assigns the appropriate weights to fires (as well as NDVI trends) for each subsample. This flexibility is necessary, given the varied use of fires in deforestation worldwide. Table (2) presents the fire densities in deforestation "hotspots". It is clear that as we begin to process fires in Madagascar or the Democratic Republic of the Congo, we will have to rely on more information to appropriately filter relevant fires. However, for the Brazilian and Indonesian subsamples, we find a high level of association between fires and deforestation. The fit can only be improved by more detailed information on fire characteristics.

3.3. **Deforestation Probability Model Specification.** We do not expect the empirical link between observed deforestation and our indicators to be perfect in all cases. First, there will inevitably be errors in the signals provided for individual pixels by remotely sensed data on fires and changes in vegetation color. There will be some false positives (spuriously-reported fires and/or NDVI changes) and false negatives

---

[13]With greatly appreciated assistance from our colleague David Roodman, we have coded the regression algorithm for Equation (1) in Mata, Stata's matrix programming language. The computation time saved is extraordinary: For one million pixels, a full set of monthly regressions runs in approximately 10 hours in Stata, and in less than 5 minutes using Mata. Similar speeds should be achievable using Matlab or the Python NumPy library.

[14]In a typical Brazilian pattern of deforestation, trees are cleared (perhaps using fires) at the beginning of the dry season. The felled trees are cleared at the end of the dry season using large-scale fires, once the timber has dried [11].
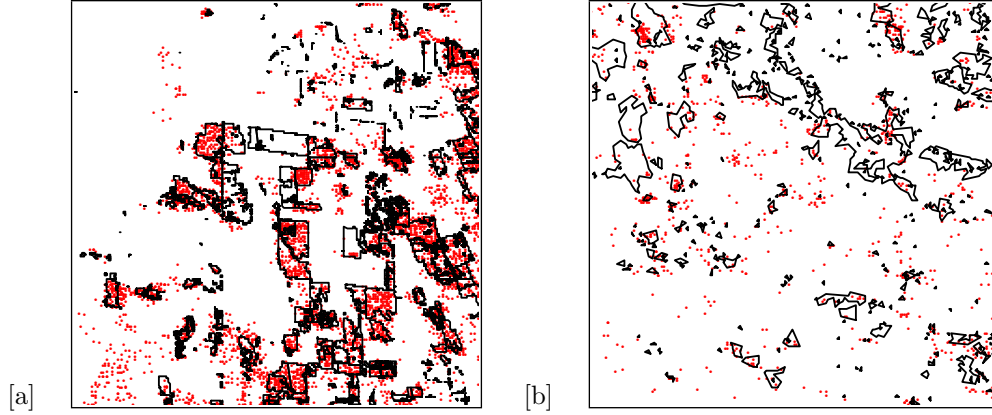
FIGURE 4. Fires and deforestation [a] PRODES (black outlines) and fires (red points), 2000-2005, for Pará subsample [b] FCLH (black outlines) and fires (red points), 2000-2005, for Riau subsample.

TABLE 2. Fire Density, Jan 2000 - Mar 2009

Country	Fires	Land Area (km^2)	Fires per km^2
Brazil	2,816,688	8,456,510	0.333
Indonesia	382,666	1,826,440	0.210
Madagascar	371,240	581,540	0.638
Bolivia	443,819	1,084,390	0.409
DRC	2,578,292	2,267,600	1.137

(unreported fires and/or NDVI changes). Since observations for fires and NDVI reflect separate phenomena, we expect to gain robustness by using both variables.[15] Second, there will be some errors in our calibration datasets: the identification of FCLH deforestation (automated) and PRODES (human interpretation) will inevitably include some false positives and false negatives.

In the same vein, the requirement that an area be labeled "deforested" or not in the FCLH and PRODES datasets forces an interpretive judgment into a statement about a discrete outcome. In fact, all interpretations without rigorous ground-truthing are statements about probability, at least implicitly. For these reasons, we believe that a probability model is most appropriate in this context. The model fits a variable coded 1 (deforestation) or 0 (not) to right hand variables derived from the MODIS data. These variables incorporate information on fires and changes in NDVI, in both the pixel itself and neighboring pixels. Formally, we specify the probability model for an observation period as follows:

$$
(3) \qquad P\{d_i = 1 | \mathbf{x}\} = \Phi(\beta_0 + \beta_1 \widehat{\alpha}_{1i} + \beta_2 \widehat{t\alpha}_{1i} + \beta_3 maxdrop_i + \beta_4 fires_i +
$$
$$
+ \beta_5 \widehat{n\alpha}_{1i} + \beta_6 \widehat{nt\alpha}_{1i} + \beta_7 nmaxdrop_i + \beta_8 nfires_i + \gamma Int'_i + \epsilon_i)
$$

---

[15]Since both variables are drawn from MODIS data, some common error components may be introduced by randomly-occurring technical problems in the MODIS system.

The variables are defined by:

d_i	$=$	Deforestation in pixel i: 1 if deforestation within pixel, 0 otherwise.
$\widehat{\alpha}_{1i}$	$=$	Estimated long-term NDVI change in pixel i from Equation (1)
$\widehat{t\alpha}_{1i}$	$=$	Estimated t-stat associated with $\widehat{\alpha}_{1i}$ from Equation (1)
$maxdrop_i$	$=$	Short-term NDVI drop, from Equation (2)
$fires_i$	$=$	Log[1 + (total fires in pixel i)]
$\widehat{n\alpha}_{1i}$	$=$	Average of $\widehat{\alpha}_{1i}$ in neighbors of pixel i
$\widehat{nt\alpha}_{1i}$	$=$	Average of $\widehat{t\alpha}_{1i}$ in neighbors of pixel i
$nmaxdrop_i$	$=$	Average of $maxdrop_i$ in neighbors of pixel i
$nfires_i$	$=$	Log[1 + (total fires in neighbors of pixel i)]
Int_i	$=$	Row vector of various interactions, specified in Table (3)
$\Phi(\cdot)$	$=$	Standard Normal Cumulative Distribution Function (cdf)

Expected Signs (without interaction variables): $\beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_7 < 0$; $\beta_4, \beta_8 > 0$

In words, Equation (3) associates the probability of deforestation in pixel i with a linear combination of signals and their interactions, passed through the standard normal cdf ($\epsilon_i \sim N(0, \sigma)$). Neighbors are defined as the 8 pixels adjacent to pixel i.[16] Absent any interactions between variables, we expect the probability of deforestation to increase as NDVI decreases and total fires increase in the pixel and its neighbors. In the presence of multiple interactions, however, the interpretation of the coefficients is less straightforward. The direction of the coefficients can still match the expected signs at their mean values, but this becomes more difficult to see in a simple tabulation of results. Multiple specifications of the model are presented in Table (3), although only the best-fitting specification is used to generate the probability maps.

In sum, the probability model allows us to relate ecological pressure to deforestation. The explanatory variables in Equation (3) tag ecological shifts, which are sometimes very rapid (e.g., persistent fires within a pixel or steep and persistent drops in the NDVI). The probit (standard normal) specification allows us to relate the signals from the explanatory variables to the best available data on actual deforestation. In this sense, our probit specification "trains" FORMA to interpret monthly signals from fires and vegetation color changes, based on the local historical relationships between these proximate determinants of deforestation and observed deforestation. Because such relationships may be quite different across areas with different ecological, economic and social dynamics, we do not expect FORMA's parameters to be the same in different regions. For this reason, we re-estimate the parameters for each 100 km^2 territorial block, and are experimenting with estimating the model for individual ecoregions.[17] We are also investigating generating monthly or bi-monthly probability maps using MODIS data available at higher spatial and temporal resolution, such as the 250 m or 500 m 16-day NDVI datasets. Data quality provides the only inherent obstacle to improvements in spatial and temporal resolution, since problems due to mechanical failure, persistent cloudiness and other factors have progressively greater impacts as the reporting interval shortens.

3.4. **Model Results for Brazil and Indonesia.** We test the model for 100 km^2 territorial blocks in Pará State, Brazil and Riau Province (Sumatra), Indonesia. In each case, we estimate the model by probit, with 10,000 observations. Table (3) reports parameter estimates and t-statistics for the linear and interactive

---

[16]The definition of neighboring pixels can be extended to the "second ring" about pixel i (16 more pixels), or even further. By considering only the pixels with centroids within 1.4km of the centroid of pixel i – the first-ring neighbors – we make an implicit assumption about the strength of the spatial autoregressive parameter. We are currently experimenting with other definitions of neighboring pixels, including information on directionality that incorporates information on which pixel "transmitted" deforestation to a given pixel (i.e. the "origin" of the proximate determinants of deforestation).

[17]Expanding the sample size for the probability model yields more degrees of freedom for estimation, but it may also degrade the potential fit for separate biomes within the estimation area. To guard against this version of the modifiable areal unit problem [3], we will ultimately confine estimation to areas within individual ecoregions identified by WWF International. We expect that the biological and climatological characteristics of an ecoregion are on average more homogeneous than those of an arbitrarily defined block of pixels – for example, a 100 km^2 subsample.

<div align="center">Table 3. FORMA Results</div>

	Brazil		Indonesia	
	(1)	(2)	(3)	(4)
$\widehat{\alpha}_1$	-0.0984***	0.0592**	0.00975	-0.0265*
	(-8.128)	(2.540)	(1.097)	(-1.861)
$\widehat{t\alpha}_1$	-0.103**	-0.323***	-0.290***	-0.0824
	(-2.472)	(-6.001)	(-5.340)	(-1.193)
$maxdrop$	0.000353	0.00478***	-0.00451***	-0.00414***
	(0.492)	(4.956)	(-11.32)	(-9.887)
$fires$	0.332***	0.228	0.388***	0.324
	(4.206)	(1.430)	(3.990)	(1.254)
$\widehat{n\alpha}_1$	-0.152***	-0.216***	-0.0189	-0.0198
	(-7.917)	(-9.333)	(-1.423)	(-1.474)
$\widehat{nt\alpha}_1$	0.273***	0.328***	0.263***	0.270***
	(3.911)	(4.275)	(3.190)	(3.257)
$nmaxdrop$	-0.00219**	-0.00144	-0.0163***	-0.0158***
	(-2.002)	(-1.221)	(-25.26)	(-23.86)
$nfires$	0.681***	1.762***	1.976***	1.788***
	(5.597)	(8.869)	(13.14)	(11.42)
$fires * \widehat{\alpha}_1$		0.00261		-0.00546
		(0.211)		(-0.908)
$nfires * \widehat{n\alpha}_1$		0.116***		-0.0325***
		(6.103)		(-2.691)
$\text{D}[\widehat{t\alpha}_1 < -3.28] * \widehat{\alpha}_1$		0.00198		-0.0398***
		(0.269)		(-9.104)
$\text{D}[\widehat{nt\alpha}_1 < -3.28] * \widehat{n\alpha}_1$		0.00197		-0.0138**
		(0.255)		(-2.068)
$maxdrop * \widehat{\alpha}_1$		0.000821***		-9.05e-05***
		(7.858)		(-3.170)
$maxdrop * fires$		-0.00127		-3.62e-05
		(-1.065)		(-0.0225)
Observations	10000	10000	10000	10000
pseudo-R^2	0.7263	0.7395	0.3649	0.3791

Note: The expression D[**x**] evaluates to 1 if **x** is true, and 0 otherwise.

Variable definitions found in Section 3.3.

*** p<0.01, ** p<0.05, * p<0.1

z statistics in parentheses

specifications. Columns (1) and (3) report linear results for Brazil and Indonesia, respectively. In both cases, we find highly significant pixel-specific and neighborhood effects for change in NDVI, significance of NDVI change, and fires. As expected, the pattern of significance for NDVI is different in the two country cases. The pixel and neighborhood results are very strong for the long-run time-trend measures ($\widehat{\alpha}_1$, $\widehat{n\alpha}_1$) in Brazil, and for the short-term drop measures ($maxdrop$, $nmaxdrop$) in Indonesia. As we noted previously, this difference reflects differences in ecological succession after deforestation. Although we gain explanatory power by including both NDVI time trends ($\widehat{\alpha}_1$, $\widehat{n\alpha}_1$) and their significance ($\widehat{t\alpha}_1$, $\widehat{nt\alpha}_1$), the four variables
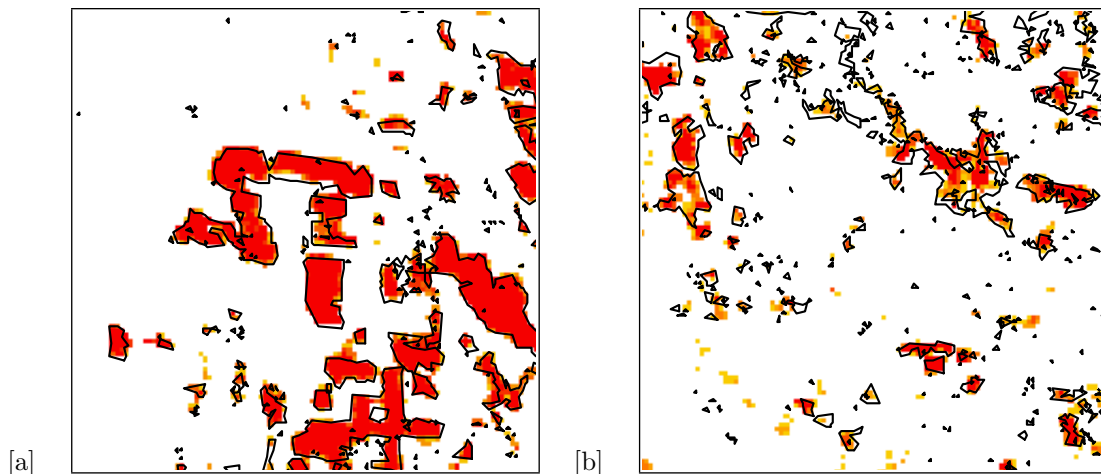
FIGURE 5. Results (spatial fit) of FORMA with FCLH deforestation for 2000-2005, [a] Pará subsample and [b] Riau subsample.

are highly collinear. When we estimate the equations for trends or significance separately, the estimated parameters have the expected signs and high significance. When we incorporate both variables, $\widehat{\alpha}_1$, $\widehat{n\alpha}_1$, and $\widehat{t\alpha}_1$, all have the expected signs and high significance. However, collinearity results in sign reversal for $\widehat{nt\alpha}_1$,, the neighborhood significance measure.

Columns (2) and (4) extend the specifications to interactions among fires, NDVI change and significance. Individual coefficients are more difficult to interpret in the interactive specifications, but we employ them for FORMA because the interactions clearly improve the fit to observed deforestation. Both linear and interactive specifications yield pronounced differences in estimated parameters between the Pará and Riau subsamples. This reflects our prior expectations, since the biological, economic and social dynamics of the two areas are very different.

Most importantly, from our applied perspective, identification (or "prediction") of deforestation and non-deforestation with this probability model is highly accurate in both sample areas. Table (4) (p. 13) tabulates FORMA-predicted probabilities against FCLH deforestation from Hansen, et al. (2008) for all 10,000 pixels in the Pará and Riau subsamples, respectively. The strength of the results reflects the very high significance of the estimated parameters. In both cases, extremely high percentages of pixels with predicted deforestation probabilities above 0.90 are identified by the FCLH map as deforested. In the Brazilian case, FORMA agrees with FCLH on deforestation for 100% of the 227 category 1.00 pixels and 98.7% of the 1,048 category [0.90-1.00] pixels.[18] In Indonesia, FORMA agrees with FCLH for 96.8% of the 349 category [0.90-1.00] pixels.

Agreement is also extremely close for non-deforestation. In Brazil, FORMA agrees with FCLH on non-deforestation for 98.9% of the 6,974 pixels in category (0.00-0.10). In Indonesia, FORMA agrees with FCLH for 96.3% of 5,989 pixels in category (0.00-0.10). Over the intermediate ranges, the percentage agreement corresponds closely to FORMA-assigned probability ranges. In Brazil, for example, FCLH and FORMA agree on deforestation for 36.6% of cases in FORMA's probability category [0.30-0.40].

Although agreement at the tails is quite close in both cases, the intermediate-range fit for Brazil is much better than for Indonesia. In Brazil, over 80% of pixels with FCLH deforestation have an estimated FORMA probability of 50% or greater. For the Brazilian subsample, we are identifying a great majority of the deforested areas using only NDVI and fire signals. In Indonesia, by contrast, only 40% of the FCLH-identified deforested pixels have an estimated FORMA probability of 50% or greater. To clarify

---

[18]The category 1.00 pixels are machine rounded to "1" due to a limited number of available digits. That is, the probability of deforestation is extraordinarily high.

TABLE 4. Tabulation of FORMA Results

Estimated Probability Range	Brazil					Indonesia				
	$d_i = 0$		$d_i = 1$		Total	$d_i = 0$		$d_i = 1$		Total
	No.	%	No.	%	No.	No.	%	No.	%	No.
$(0.00 - 0.10)$	6,894	98.9	80	1.1	6,974	5,769	96.3	220	3.7	5,989
$[0.10 - 0.20)$	416	86.7	64	13.3	480	1,339	86.3	213	13.7	1,552
$[0.20 - 0.30)$	202	75.9	64	24.1	266	544	75.3	178	24.7	722
$[0.30 - 0.40)$	123	63.4	71	36.6	194	281	66.4	142	33.6	423
$[0.40 - 0.50)$	81	52.3	74	47.7	155	159	54.1	135	45.9	294
$[0.50 - 0.60)$	72	52.6	65	47.4	137	93	41.2	133	58.8	226
$[0.60 - 0.70)$	50	36.0	89	64.0	139	49	28.0	126	72.0	175
$[0.70 - 0.80)$	31	19.5	128	80.5	159	35	24.8	106	75.2	141
$[0.80 - 0.90)$	21	12.3	150	87.7	171	17	13.2	112	86.8	129
$[0.90 - 1.00)$	14	1.3	1,034	98.7	1,048	11	3.2	338	96.8	349
1.00^*	0	0	277	100	277	0	0	0	0	0
Total	7,904	79	2,096	21	10,000	8,297	83	1,703	17	10,000

*Note that a probability of exactly one indicates a number $(0.99\overline{9})$ beyond machine precision.

further, FORMA does not have problems with false positives in Indonesia, since almost all of the highest-probability FORMA pixels are identified as deforested by the FCLH map. There are problems, however, with false negatives in Indonesia, since nearly 60% of FCLH-identified pixels do not have FORMA-estimated probabilities over 50%.

Table (5) (p. 14) suggests that the high false-negative incidence for Indonesia may be caused by our use of relatively coarse-resolution NDVI data, which have been resampled to 1 km from 250 m data. Justice, et al. (1989) suggest that resampling to a lower resolution often dilutes satellite signals. FCLH deforestation is reported at 500 m resolution, while we use signals with 1 km resolution. In our current analysis, we identify a pixel as having experienced deforestation if FCLH shows deforestation in any one of the four 500 m pixels that fall within it. In an auxiliary analysis, we found that FORMA's results for Indonesia include far fewer false negatives for pixels where three or four of the 500 m FCLH pixels are tagged with deforestation. Most of the problem stems from failure to identify deforestation in pixels where only one or two 500 m pixels are affected. Therefore, the difference in false-negative results for Brazil and Indonesia is apparently due to the presence of far more small, isolated patches of deforestation in Indonesia.

We tentatively conclude that the incidence of false-negative results in Indonesia could be significantly reduced by using existing NDVI data at higher resolution. This would also require moving from 32-day to 16-day composites,[19] but would enable us to estimate the probability model for individual FCLH pixels at 500 m, allowing us to focus more precisely on deforestation in scattered, isolated pixels.[20] The potential gain seems to warrant the additional computational burden, and we will make this adjustment for the next version of FORMA.

Here we should emphasize that FORMA has negligible problems with false positives, so the basic power of the FORMA filter is not diminished. Nearly all of the main concentrations of deforestation are tagged appropriately in both the Brazilian and the Indonesian subsamples. Only some of the fringe or peripheral deforestation pixels are missed. The results are illustrated by Figures (5a) and (5b) (p. 12), which display the spatial match between FCLH deforestation and FORMA-predicted deforestation for the subsamples in Pará, Brazil and Riau, Indonesia. In the figures, FCLH-identified pixels are outlined in black. We employ a cutoff of 0.50 for the FORMA predictions and don't color any pixels with predicted probabilities below that

---

[19]The NDVI at 1 km resolution is available in 32-day composites. At higher spatial resolutions, only 16-day composites are available. This higher temporal resolution may necessitate the adjustment of certain parameters, but nothing substantive.

[20]The current definition of relevant "neighbors" makes certain implicit assumptions about the spatial autoregressive parameter, which depends on the size of the MODIS pixel. If we move to a higher resolution, we may have to include more neighborhood "rings" of pixels, depending on the relative effects of higher spatial *and* temporal resolution.

TABLE 5. Average Intensity of FCLH Deforestation, by FORMA probability

	Brazil	Indonesia
Average FCLH "hits" per pixel		
$0.0 < \mathrm{P}\{d_i = 1\} \leq 0.1$	1.288	1.355
$0.1 < \mathrm{P}\{d_i = 1\} \leq 0.2$	1.453	1.676
$0.2 < \mathrm{P}\{d_i = 1\} \leq 0.3$	1.703	2.135
$0.3 < \mathrm{P}\{d_i = 1\} \leq 0.4$	1.676	2.127
$0.4 < \mathrm{P}\{d_i = 1\} \leq 0.5$	1.959	2.378
$0.5 < \mathrm{P}\{d_i = 1\} \leq 0.6$	2.031	2.541
$0.6 < \mathrm{P}\{d_i = 1\} \leq 0.7$	2.180	3.032
$0.7 < \mathrm{P}\{d_i = 1\} \leq 0.8$	2.359	2.925
$0.8 < \mathrm{P}\{d_i = 1\} \leq 0.9$	2.633	3.321
$0.9 < \mathrm{P}\{d_i = 1\} \leq 1.0$	3.712	3.538
No. of pixels (of the 10,000 total)		
FCLH = 0	7904	8297
FCLH = 1	344	545
FCLH = 2	319	336
FCLH = 3	255	249
FCLH = 4	1178	573

cutoff. Higher probabilities are presented as shades of yellow to red representing increments of 0.10, with dark red showing $P\{d_i = 1\} \geq 0.90$. In both cases, the extremely close alignment with FCLH-identified pixels is striking. The original motivation for FORMA – a rapid-update system to flag deforestation – is clearly fulfilled for both the Brazilian and Indonesian subsamples, even if some of the peripheral or far-flung deforested pixels are missed (due in part to the coarse spatial resolution of this first-generation model, as we noted previously).

In this context, we should note the importance of neighborhood effects in improving the visual display of FORMA results. To illustrate, Figures (6a) and (6c) present FORMA results without and with neighborhood effects for the Pará subsample; Figures (6b) and (6d) present results without and with neighborhood effects for the Riau subsample. The panels in the top and middle rows include the full range of color-coded probabilities, while the panels in the bottom row provide color-coded probabilities that are 0.50 or greater. The effect of spatial smoothing is the elimination of significant "clutter" in the top-row panels caused by false-positives from isolated fires or NDVI drops. When neighborhood effects are included, estimated probabilities for isolated pixels are usually driven below the cutoff display level (bottom-row panels). This process can be thought of in terms of a topographic map, where we first smooth isolated spikes, and then select pixels that are above a certain altitude (0.50 or greater).

The inclusion of neighborhood effects clearly improves the interpretive capabilities of the FORMA filter. The displayed probability for the Pará subsample [Figure (6e)] is compared with both FCLH and PRODES in Figure (7) (p. 16). The spatial smoothing highlights (appropriately) the broad swaths of deforestation that are picked up by both FCLH and PRODES. FORMA assigns a greater probability of deforestation if a pixel is adjacent to pixels with strong signals (multiple fires or significant NDVI drops). This suggests that there may be a critical mass element to the identification of deforestation: For a set of contiguous pixels, as the signals become stronger within each pixel, the probability of deforestation will rise for the entire set. In Figure (7d), we juxtapose our FORMA/FCLH overlay for Pará with deforested pixels identified by PRODES. Since PRODES is operated specifically for Brazil by Brazilian scientists, we expect it to provide the most accurate representation of deforestation in Pará. The comparison yields two particularly noteworthy results. The first, which is important in the global context, is that FCLH-identified pixels line up exceptionally well with PRODES-identified pixels. This provides reassuring support for using FCLH deforestation estimates to calibrate FORMA outside of Brazil. The second is that FORMA-predicted pixels line up remarkably well with PRODES, as well as with FCLH, which is encouraging for our plans to apply FORMA globally.
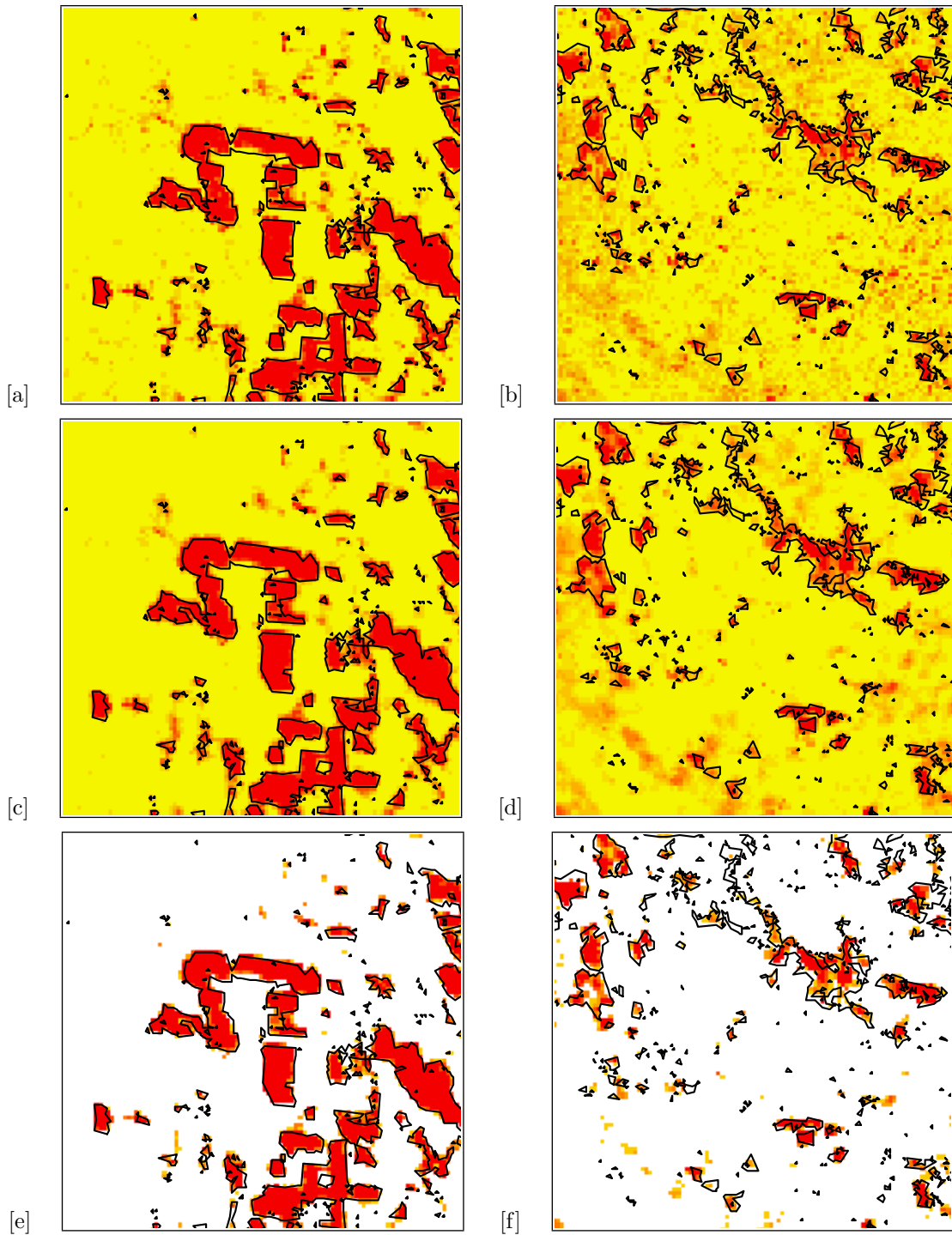
FIGURE 6. Model Results, with and without neighborhood effects (2000-2005): [a] and [b] model specified without neighborhood characteristics (spatial smoothing) for Pará subsample and Riau subsample, respectively; [c] and [d] model with spatial smoothing, nearest eight pixels (first ring); [e] and [f] only estimated probabilities greater than 0.50 colored. FCLH deforestation (2000-5) in black outlines.
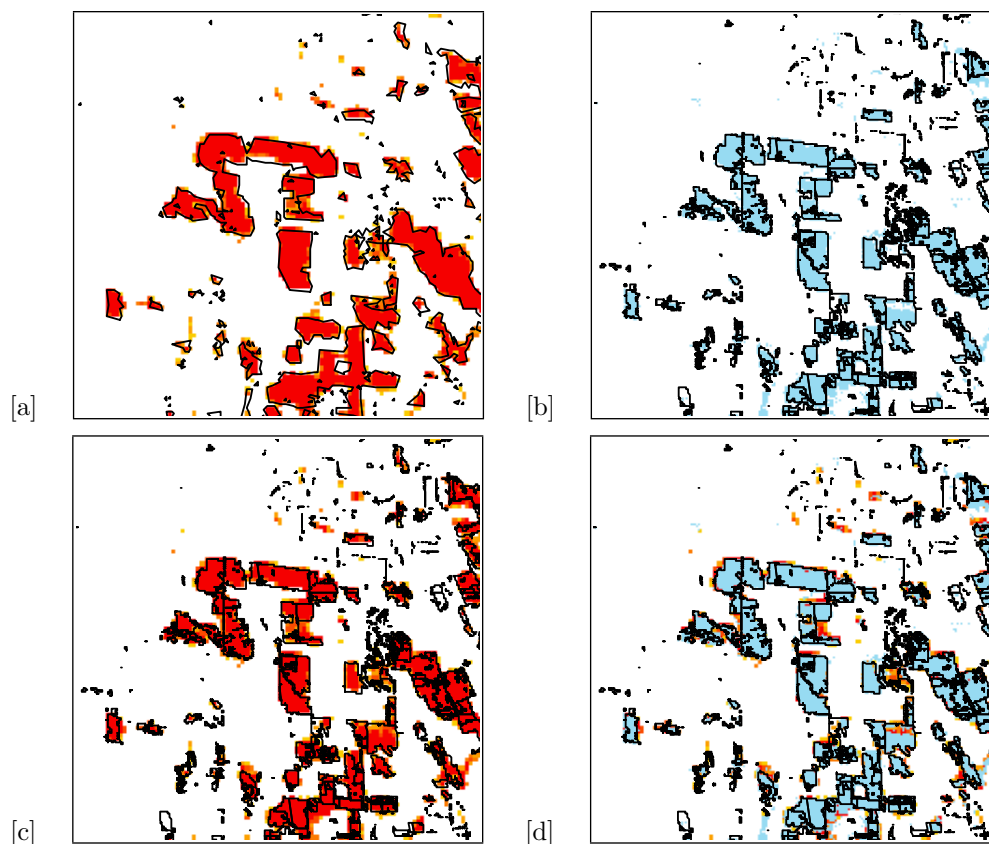
FIGURE 7. Comparison of deforestation estimates. [a] FCLH (black outlines) and FORMA probabilities (in red) [b] PRODES (black outlines) and Hansen (in blue) [c] PRODES (black outlines) and FORMA probabilities (in red) [d] PRODES (black outlines) and FORMA probabilities (in red) underneath FCLH (in blue).

4. A DYNAMIC TEST FOR PARÁ, BRAZIL

The ultimate test of FORMA's relevance for global monitoring is its ability to capture short-term changes in deforestation. We are able to test this ability in Pará, because PRODES has identified new deforestation in our test-area pixels for the years 2006, 2007 and 2008. This provides an excellent test for FORMA, which is fitted to FCLH deforestation data for the period 2000-2005. We begin with the FORMA parameters estimated from FCLH data for the Pará test block during 2000-2005 [Table(3)]. We construct full datasets for our MODIS variables for the periods 2000-2006, 2000-2007 and 2000-2008. We apply the estimated parameters to these datasets, yielding deforestation predictions for all 10,000 pixels for the three periods. Then, successively for 2006, 2007 and 2008, we identify pixels whose predicted deforestation probabilities exceed 0.50 for the first time in that year. Figure (8) (p. 17) overlays these predictions with newly-identified deforestation in PRODES for the same year. The overall match from year to year is strikingly close, and, as Figure (8d) and Figure (9d) show, the match for the period 2006-2008 is nearly exact. Nearly all of the newly-identified pixels accurately tag areas of PRODES deforestation over the previous year; most of the remaining pixels are at the periphery of deforested tracts from the prior year.

By the same logic we proposed in the earlier comparison with FCLH-identified pixels, we believe that some of the FORMA "false positives" and "false negatives" underlying the results in Table (4) (p. 13) may not actually be false. We have three reasons for suspecting this to be the case. First, imagery for the PRODES deforestation maps is acquired in August of each year to produce maps for the previous 12 months, so a
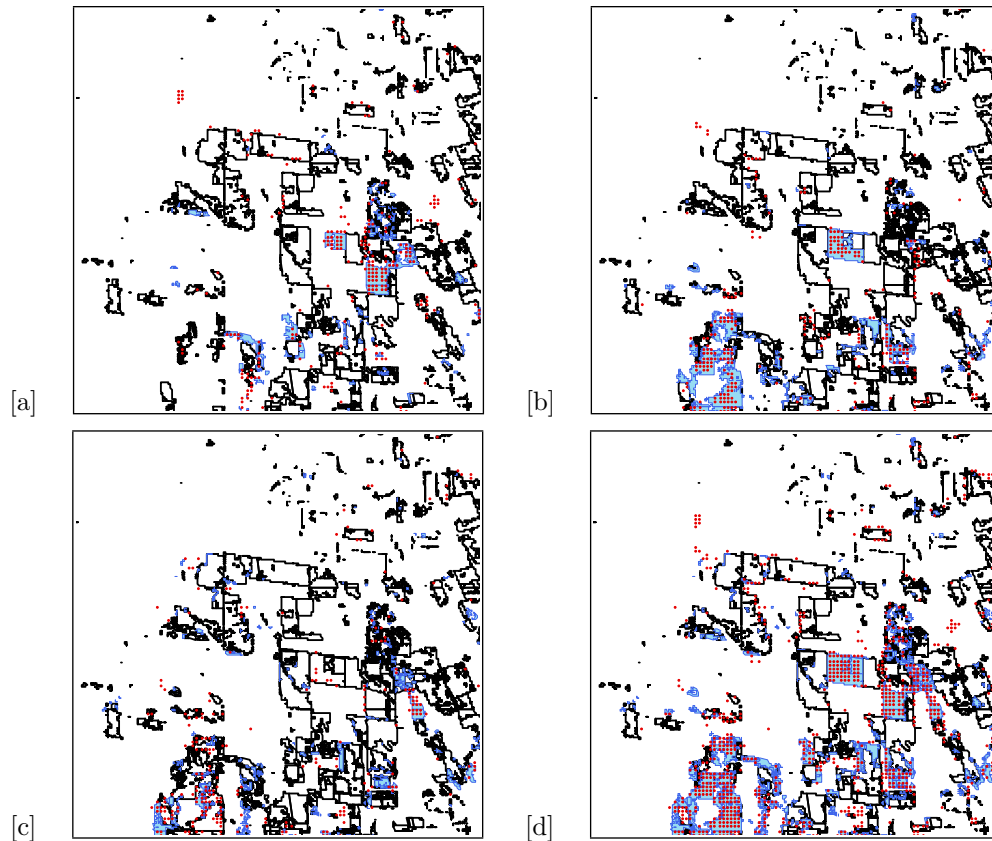
FIGURE 8. Newly Identified PRODES and FORMA deforestation, 2006-2008. [a] New deforestation in 2006: PRODES (blue polygons) and FORMA (red points), with historical PRODES deforestation (gray outlines) [b] New deforestation in 2007: PRODES (blue) and FORMA (red) [c] New deforestation in 2008: PRODES (blue) and FORMA (red) [d] cumulative PRODES (blue) and FORMA (red) registering between 2006 and 2008.

"PRODES year" is not quite the same thing as a "FORMA year". This temporal misalignment causes some false hits that are resolved in the next period. Second, and related to this, the three-year alignment with PRODES is near-exact [see Figure (9d)], so we are reluctant to label interim deviations as false positives or negatives. Third, we re-invoke our original, common-sense proposition about FORMA's ability to explain deforestation. Using MODIS information that is not identical to the information used by PRODES, FORMA identifies high probabilities of deforestation in areas that have pronounced long- and/or short-term NDVI decreases, a high incidence of fires, and strong neighborhood effects. In the comparison with PRODES, FORMA "false positives" are pixels that PRODES identifies as not deforested, despite the presence of large drops in NDVI, repeated fires, and strong neighborhood effects. In the same vein, FORMA "false negatives" are pixels that PRODES identifies as deforested, despite weak or absent drops in NDVI, few or no fires, and weak or nonexistent neighborhood effects. Given that PRODES itself depends on remotely sensed data, we find it unlikely that all of the "false" reads in this context are in FORMA.
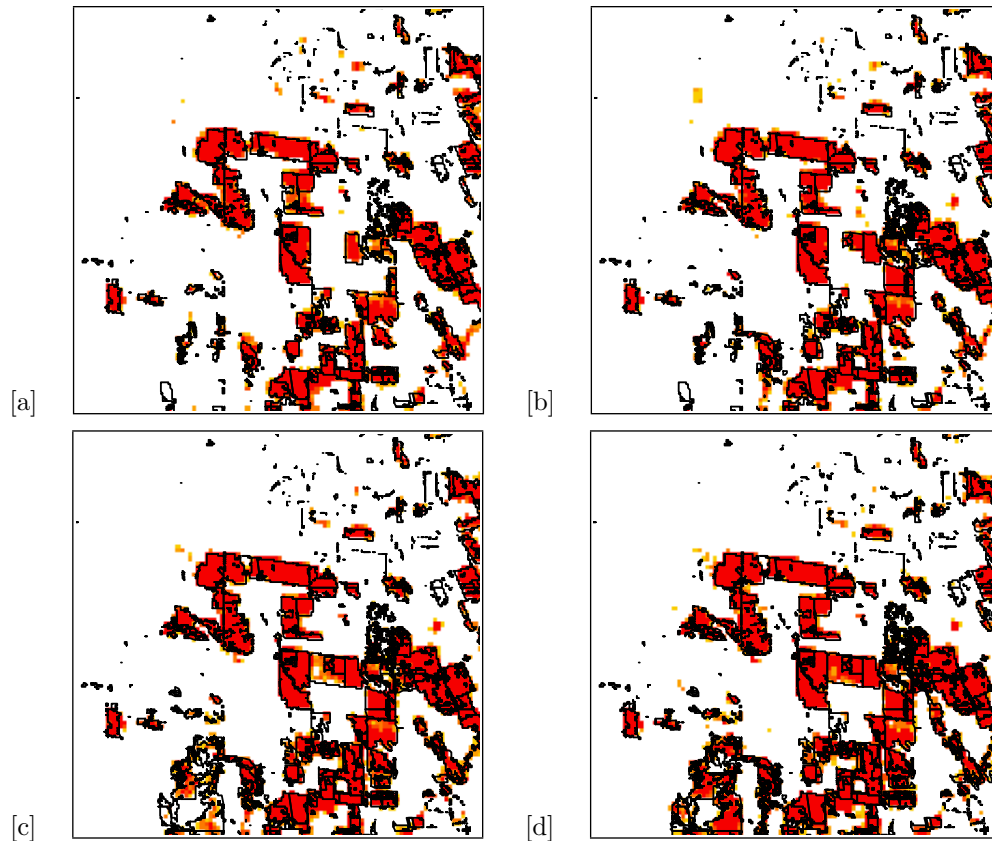
FIGURE 9. PRODES and FORMA deforestation (2006-2008): [a] PRODES (black outline) and FORMA (in red), cumulative deforestation estimates for 2000-2005 [b] PRODES and FORMA, cumulative deforestation estimates for 2000-2006 [c] PRODES and FORMA, cumulative deforestation estimates for 2000-2007 [d] PRODES and FORMA, cumulative deforestation estimates for 2000-2008.

5. Quarterly Deforestation Dynamics: An Illustration for Riau

To illustrate FORMA's potential for rapid identification of deforestation patterns, we develop quarterly predictions of deforestation for the Riau subsample from Q1 2006 to Q4 2008.[21] Figure (10) (p. 19) overlays areas with high predicted deforestation (colored red) at 3-month intervals on a map of Riau that includes protected areas (colored gray).[22,23] These areas are among the last refuge zones for the critically-endangered Sumatran tiger. Although we cannot test the FORMA predictions against actual deforestation, the successive extensions of predicted deforestation from previously-deforested areas look quite plausible. Little deforestation is visible within the protected areas in Q1 2006, but it begins spreading early in the series, and then invades at an alarming rate. Hansen et al. (2009) note a marked increase in deforestation in Indonesia from 2000 to 2005, and our Riau exercise may illustrate its extension in this sample area through

---

[21]We predict the probability of deforestation in each cell in each quarter using all available time series observations for fires and NDVI in that cell and its neighbors.

[22]The output probabilities are generated every month, but without monthly calibration data, there is the possibility of spurious high probabilities that are quickly reduced before the next month's probability output. We employ a three-month moving average for smoothing the final probabilities to produce quarterly (rather than monthly) estimates. We are currently running data checks to see if the rolling time series block length can be reduced even further. Without contemporaneous calibration data, the checks will have to be internal – but the checks are possible nonetheless.

[23]A PDF animation of the quarterly changes is available at http://www.cgdev.org/forest.
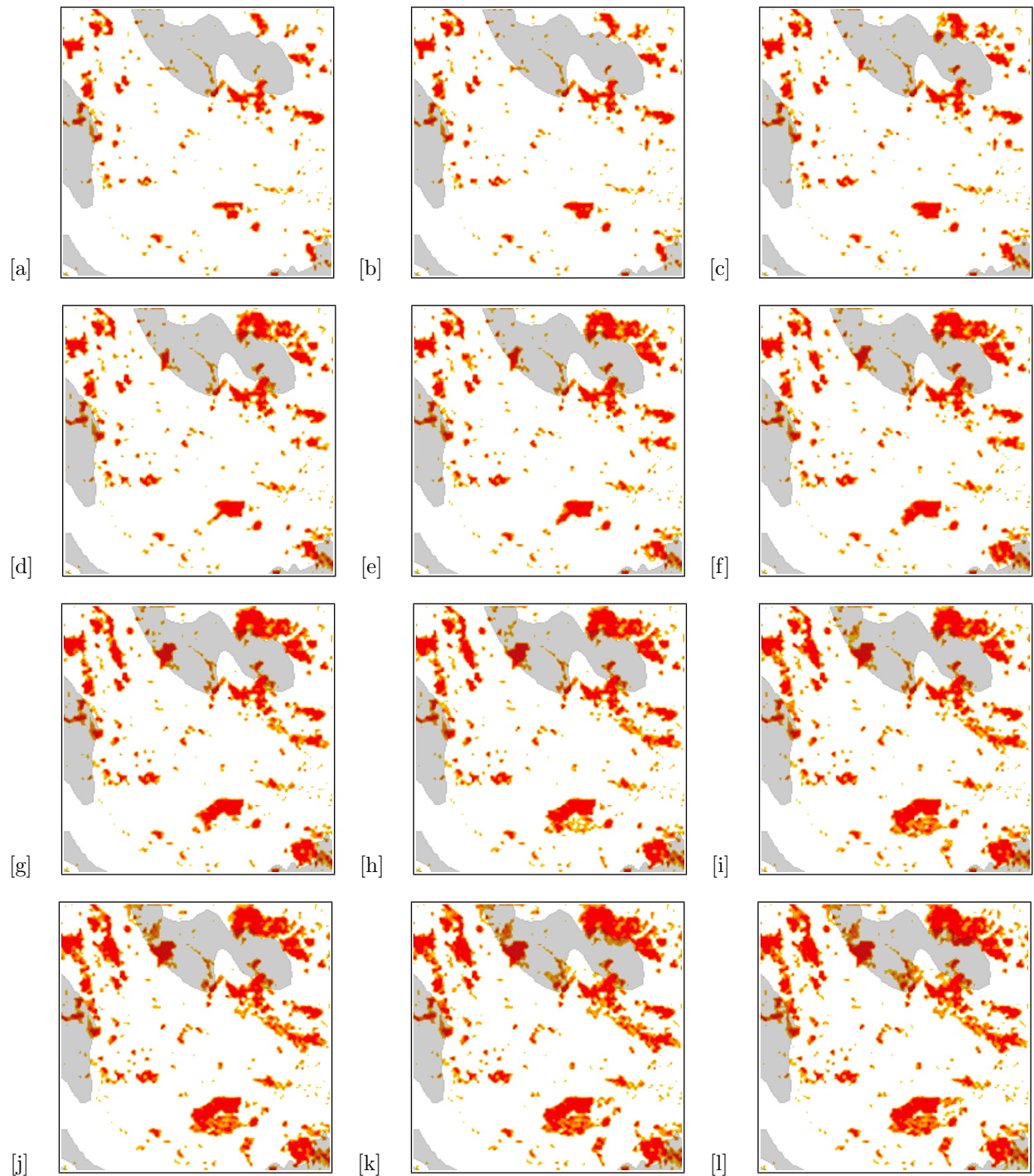
FIGURE 10. FORMA output for Riau subsample, Q1 2006 - Q4 2008 (left to right, top to bottom). The output is colored yellow to red according to increasing probability of deforestation as of the specified quarter. The difference between each map and its predecessor represents probable deforestation *within* the specified quarter. The output is spline interpolated for ease in visualization and interpretation. The transparent gray polygons show the boundaries of protected areas.

the rest of the decade. Figure (10) (p. 19) demonstrates the potential power of a global system that can rapidly identify emerging deforestation hotspots. In Riau, such information could have focused Indonesian and international attention on rapid destruction of the Sumatran tiger's habitat.

6. Future Extensions

6.1. **Geographic Expansion.** During the next few months, we will expand FORMA toward full coverage of the tropical moist forest biome. We have coded the system for speed, so we do not anticipate significant computational constraints. Our first objective is full coverage for Indonesia, along with completion of Brazilian coverage to permit further refinement and testing against PRODES. Central Africa and Madagascar will be next, followed by expansion to complete coverage of the remaining moist tropical forest areas of Asia, Africa and Latin America.

6.2. **Improved Calibration Data.** The main calibration dataset for FORMA is FCLH (from Hansen, et al. (2008)), but as in Brazil, where PRODES data are available, FORMA can take advantage of improved calibration datasets without waiting for the latest state-of-the-art, pan-tropical dataset. For example, if more accurate, higher resolution deforestation maps become available through UN-REDD assistance or domestic mapping programs, one calibration dataset can be swapped for another, improving FORMA estimates for that particular subregion of the tropics. FORMA also depends on forest cover maps for 2000 and 2005. More accurate and higher-resolution forest cover maps would also improve the accuracy of FORMA by limiting analysis to forested areas exclusively.

6.3. **Dissemination with Google Earth and Google Maps.** In keeping with our open-source commitment, FORMA's deforestation predictions will be developed as overlays for Google Earth and Google Maps. We believe that this approach will maximize FORMA's potential impact in two ways. First, it will make the outputs available to concerned global citizens, regardless of their technical expertise. Second, using Google Earth and Google Maps will enable us to open the system to direct ground-truthing, via GPS-identified photographs and text commentaries uploaded by interested stakeholders with direct access to the reporting areas. This feedback will improve FORMA, and move local evidence on deforestation into the international arena.

Figure (11) illustrates our current thinking about integration with Google Earth, using a display that shows precisely why a particular area is tagged as deforested.[24] Probable deforestation appears as a raster image overlay that is transformed into color-coded placemarks as the user zooms in. The standard color scheme applies, presenting probabilities from .50 to 1 as shades of yellow to red. Clicking on a placemark reveals a complete information set for the associated pixel, including its current FORMA deforestation probability, location information, dynamically generated time series graphs for NDVI and fires (including the period when critical signals were detected), and an area for links to GPS-located photos and text comments.

6.4. **The Potential Power of Timely Global Monitoring.** We believe that this user-friendly format will facilitate input from local stakeholders, which will serve several purposes. First, by revealing areas that are under particular threat, it will motivate local stakeholders to intervene more actively. Second, by rapidly disclosing deforestation in protected areas, it will replicate the role played by DETER and PRODES in Brazil: focusing immediate attention from national and local environmental regulators on the problem area. Third, by identifying land-clearing activity at a fairly high level of spatial resolution, it will facilitate the identification of actors who are clearing forests illegally. For at least some of these actors (particularly larger ones), this will bring reputational concerns to bear on continued deforestation. It may also deter land-clearing by raising the prospect of nearly-immediate discovery and global disclosure once it has begun.

All of these factors bear on our main reason for developing FORMA: its potential relevance for UN-REDD and similar programs. As we noted in the introduction, these programs are beginning to forge agreements with governments (and potentially other proprietors) for forest conservation in exchange for direct compensation. Once FORMA is fully operational, with quarterly or monthly updates, it will provide

---

[24]A preliminary Google Earth layer is available at http://www.cgdev.org/forest.
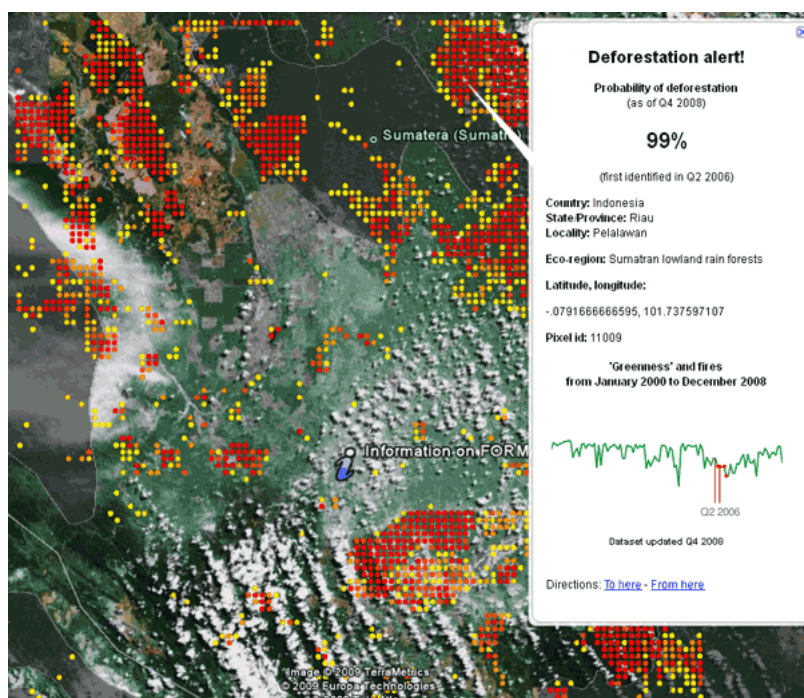
Figure 11. Google Earth Image

a continuous, neutral global monitor for such arrangements while maintaining political independence from them. We believe that this will be critical for the sustainability of UN-REDD agreements, because it will strongly diminish the contracting parties' incentives to cheat. Attention is usually focused on proprietor governments in this context, but bilateral or multilateral donor agencies are not immune from the temptation to suppress embarrassing information about failed projects. As the auditing arms of these institutions know only too well, agency professionals with career concerns are generally loath to acknowledge problems with projects that they have designed and managed. In this context, FORMA will serve internal auditors as well as global stakeholders.

6.5. **Sustainable Implementation of FORMA.** Our small team at the Center for Global Development has neither the intention nor the resources to establish a permanent home for FORMA at the Center. As we previously noted, we will pursue a completely open-source policy for the system, so that any institution or group can replicate its monitoring and reporting capabilities at modest cost. While we believe this will enhance "contestability" and keep the system honest, we recognize the need for at least one global site where FORMA, or an improved successor, is maintained and further developed. Since the only critical requirement is its independent status as a global monitor, FORMA should be housed in an institution that is insulated from the intense political pressures that are bound to emerge as billions of dollars are invested in forest conservation contracts under UN-REDD and similar programs. In principal, the system could be supported by a university, a think tank, a UNFCCC-sanctioned international agency, or a consortium of conservation NGO's. As long as it remains transparent, global scrutiny and "contestability" should suffice to sustain its reporting integrity. In any case, FORMA has been designed to operate at maximum speed, minimum cost, and, because it will be open-source, with maximum opportunity for improvement by interested participants all over the world. FORMA should be easy to sustain by two or three skilled people, working with reasonably powerful desktop PC's that have standard internet access.

7. Summary and Conclusions

This paper has described the development and testing of FORMA (Forest Monitoring for Action), a low-cost, PC-based system that uses remote sensing information from MODIS for rapid identification of new deforestation in humid tropical forest areas. Our research has focused on determining whether frequently-reported MODIS data on fires and changes in vegetation color are sufficient to provide accurate predictions of deforestation in periods as short as one quarter, and in areas as small as 1 sq km. Our results are strongly affirmative, and we have found no technical reason why FORMA's algorithms could not be applied to intervals as short as one month, in areas as small as 250 meters on a side. It is, however, possible that unavoidable "noise" in MODIS data (caused by technical problems, sporadic cloud cover, etc.) may seriously degrade the accuracy of monthly predictions. Further research will be required to determine FORMA's spatial and temporal limitations. Both are likely to ease, in any case, as remote sensing systems grow more powerful.

We have developed FORMA to address an obvious monitoring problem that threatens to undermine UN-REDD-type arrangements. In light of widespread institutional, technical and corruption problems, donor constituencies are unlikely to sustain support for UN-REDD and related programs unless they have open access to accurate, credible, timely information about the state of the forested areas that they are paying to conserve. FORMA provides a prototype for a monitoring system that meets the needed conditions. It is extremely fast, low-cost, and designed for open access and ground-truthing using free, simple online tools. In the near future, we will work with a coalition of interested stakeholders to establish permanent FORMA coverage for all humid tropical forest areas.

We are encouraged by FORMA's apparent accuracy and viability. Few doubt that forest conservation can play a major role in limiting the carbon emissions that threaten to produce a global climate catastrophe. And there can be no doubt that accurate, independent monitoring will make a critical contribution to sustainable agreements that compensate proprietors for forest conservation. Steady expansion in remote sensing capabilities has made a system like FORMA inevitable. Once open-source, accurate, global monitoring of forest conservation becomes possible, donors will be much more willing to enter into conditional agreements for forest conservation payments.

That said, we feel compelled to close on a cautionary note. In REDD-type agreements, many of the donors will be governments and bilateral or multilateral agencies, and many of their counterparts will be national, regional or local governments. But, with the advent of open, accurate monitoring, it will also become possible for private parties to enter into conservation contracts. As these opportunities emerge, the value of forested land in the tropics will rise by the discounted value of expected conservation payment flows. The system will serve the donors' intent when land ownership is clear and well-protected by legal systems. Unfortunately, these conditions do not hold in many tropical forest areas. Without close attention to safeguards, REDD-type compensation programs risk setting off violent struggles for control of newly-valuable forested land in areas with weakly-enforced tenure rights. As the developers of an information system that is intended to strengthen REDD programs, we would urge the designers of such programs to be wary of implementing them in areas where legal restraints are insufficient to prevent the strong from preying on the weak.

References

[1] Carroll, M.L., C.M. DiMiceli, J.R.G. Townshend, R.A. Sohlberg, M.C. Hansen, R.S. DeFries. 2006. Vegetative Cover Conversion MOD44A, Burned Vegetation, Collection 4, University of Maryland, College Park, Maryland.

[2] Chomitz, Ken and Andrew Nelson. (*forthcoming*). Remoteness and protection status modulate protected area effectiveness in reducing tropical deforestation: a global analysis. World Bank.

[3] Dark, Shawna and Danielle Bram. 2007. The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, 31(5)5, 471-479.

[4] FAO (UN Food and Agricultural Organization). 2008. Extent of forest and other wooded land 2005. *Available online:* http://www.fao.org/forestry/site/43035/en/

[5] Grainger, Alan. 2008. Difficulties in tracking the long-term global trend in tropical forest area. PNAS 105:818-823.

[6] Hansen, M.C., Stehman, S.V., Potapov, P.V., Loveland, T.R., Townshend, J.R.G., DeFries, R.S., Pittman, K.W., Stolle, F., Steininger, M.K., Carroll, M., Dimiceli, C. 2008. Humid tropical forest clearing from 2000

to 2005 quantified using multi-temporal and multi-resolution remotely sensed data. PNAS, 105(27), 9439-9444. www.pnas.org/cgi/doi/10.1073/pnas.0804042105

[7] Hansen, M.C., Stehman, S.V., Popatov, P.V., Arunarwati, B., Stolle, F., Pittman, K. 2009. "Quantifying changes in the rates of forest clearing in Indonesia from 1990 to 2005 using remotely sensed data sets". Environmental Research Letters.

[8] Jensen, J. R. 2007. *Remote Sensing of the Environment: An Earth Resource Perspective.* Prentice Hall: New Jersey.

[9] Justice, C.O., B. L. Markham, J. R. G. Townshend and R. L. Kennard. 1989. Spatial degradation of satellite data. International Journal of Remote Sensing, 1366-5901, 10(9), 1539 - 1561.

[10] Mayaux, Philippe, Graeme M. Buchanan, Stuart H.M. Butchart, Guy Dutson, John D.Pilgrim, Marc K. Steininger and K. David Bishop. 2008. Using remote sensing to inform conservation status assessment: Estimates of recent deforestation rates on New Britain and the impacts upon endemic birds. Biological Conservation, 141(1), 56-66.

[11] Morton, D.C., R.S. Defries, J.T. Randerson, L. Giglio, W. Schroeder and G.R. Van Der Werf. 2008. Agricultural intensification increases deforestation fire activity in Amazonia. Global Change Biology, 14, 2262-2275.

[12] Mulligan, M. 2008. RALUCIAPA : Rapid assessment of land use change in and around protected areas (2000-2005). Version 1.0. http://www.unep-wcmc.org/.

[13] PRODES. 2009. Projeto PRODES: Monitoramento da Floresta Amazonica Brasileira por Satelite. http://www.obt.inpe.br/prodes/.

[14] Stern, Nicholas, et al. 2006. Stern Review on the economics of climate change. Available here: http://www.hm-treasury.gov.uk

[15] UNFCCC (United Nations Framework Convention on Climate Change). 2007. Report on the analysis of existing and potential investment and financial flows relevant to the development of an effective and appropriate international response to climate change. Dialogue working paper 8. Vienna.

[16] World Bank. 2008. Forest Carbon Partnership Facility: a framework for piloting activities to reduce emissions from deforestation and forest degradation. Washington: World Bank.

[17] World Resources Institute (WRI). 2009. Climate Analysis Indicators Tool (CAIT). http://cait.wri.org/.

[18] World Resources Institute (WRI). 2007. Painting the Global Picture of Tree Cover Change: Tree Cover Loss in the Humid Tropics. Contributors: Stephen Adam, Susan Minnemeyer, Matthew Hansen, Peter Potapov, Kyle Pittman. http://www.wri.org/publication/painting-the-global-picture-of-tree-cover-change