

Using spatial statistics to identify emerging hot spots of forest loss

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 Environ. Res. Lett. 12 024012

(<http://iopscience.iop.org/1748-9326/12/2/024012>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 210.77.64.106

This content was downloaded on 30/03/2017 at 11:15

Please note that [terms and conditions apply](#).

You may also be interested in:

[Measurement and monitoring needs, capabilities and potential for addressing reduced emissions from deforestation and forest degradation under REDD+](#)

Scott J Goetz, Matthew Hansen, Richard A Houghton et al.

[National satellite-based humid tropical forest change assessment in Peru in support of REDD+ implementation](#)

P V Potapov, J Dempewolf, Y Talero et al.

[Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012](#)

A Tyukavina, A Baccini, M C Hansen et al.

[Humid tropical forest disturbance alerts using Landsat data](#)

Matthew C Hansen, Alexander Krylov, Alexandra Tyukavina et al.

[Multiple pathways of commodity crop expansion in tropical forest landscapes](#)

Patrick Meyfroidt, Kimberly M Carlson, Matthew E Fagan et al.

[Mapping dynamics of deforestation and forest degradation in tropical forests using radar satellite data](#)

Neha Joshi, Edward TA Mitchard, Natalia Woo et al.

[Mapping and monitoring deforestation and forest degradation in Sumatra \(Indonesia\) using Landsat time series data sets from 1990 to 2010](#)

Belinda Arunarwati Margono, Svetlana Turubanova, Ilona Zhuravleva et al.

[The future of forests and orangutans \(Pongo abelii\) in Sumatra: predicting impacts of oilpalm plantations, road construction, and mechanisms for reducing carbon emissions from deforestation](#)

David L A Gaveau, Serge Wich, Justin Epting et al.

Environmental Research Letters



LETTER

Using spatial statistics to identify emerging hot spots of forest loss

OPEN ACCESS

RECEIVED

17 August 2016

REVISED

8 December 2016

ACCEPTED FOR PUBLICATION

18 January 2017

PUBLISHED

7 February 2017

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Nancy L Harris^{1,5}, Elizabeth Goldman¹, Christopher Gabris², Jon Nordling², Susan Minnemeyer¹, Stephen Ansari², Michael Lippmann², Lauren Bennett³, Mansour Raad³, Matthew Hansen⁴ and Peter Potapov⁴

¹ World Resources Institute, 10 G Street NE Suite 800, Washington, DC, 20002, United States of America

² Blue Raster, 2200 Wilson Blvd Suite 400, Arlington, VA 22201, United States of America

³ Esri, 380 New York Street, Redlands CA 92373, United States of America

⁴ Department of Geographical Sciences, University of Maryland, College Park, MD 20742, United States of America

⁵ Author to whom any correspondence should be addressed.

E-mail: nharris@wri.org

Keywords: deforestation, spatial statistics, Brazil, Democratic Republic of Congo, Indonesia

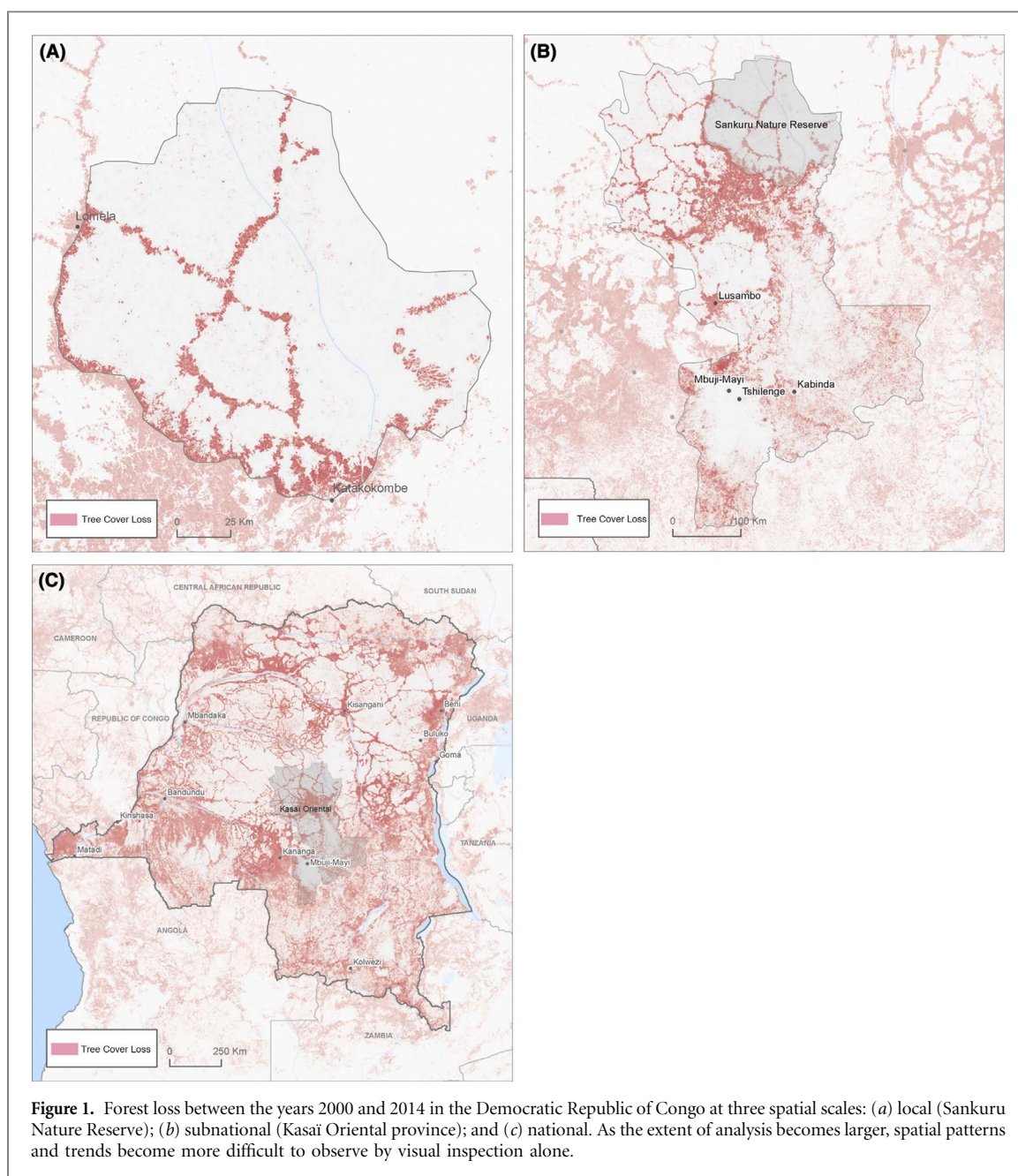
Abstract

As sources of data for global forest monitoring grow larger, more complex and numerous, data analysis and interpretation become critical bottlenecks for effectively using them to inform land use policy discussions. Here in this paper, we present a method that combines big data analytical tools with Emerging Hot Spot Analysis (ArcGIS) to identify statistically significant spatiotemporal trends of forest loss in Brazil, Indonesia and the Democratic Republic of Congo (DRC) between 2000 and 2014. Results indicate that while the overall rate of forest loss in Brazil declined over the 14-year time period, spatiotemporal patterns of loss shifted, with forest loss significantly diminishing within the Amazonian states of Mato Grosso and Rondônia and intensifying within the cerrado biome. In Indonesia, forest loss intensified in Riau province in Sumatra and in Sukamara and West Kotawaringin regencies in Central Kalimantan. Substantial portions of West Kalimantan became new and statistically significant hot spots of forest loss in the years 2013 and 2014. Similarly, vast areas of DRC emerged as significant new hot spots of forest loss, with intensified loss radiating out from city centers such as Beni and Kisangani. While our results focus on identifying significant trends at the national scale, we also demonstrate the scalability of our approach to smaller or larger regions depending on the area of interest and specific research question involved. When combined with other contextual information, these statistical data models can help isolate the most significant clusters of loss occurring over dynamic forest landscapes and provide more coherent guidance for the allocation of resources for forest monitoring and enforcement efforts.

1. Introduction

Tropical forests have undergone substantial changes over past decades, driven largely by the expansion of agricultural lands to satisfy global demand for commodities such as palm oil, timber, wood fiber, soy and beef (Ramankutty *et al* 2008). Tropical deforestation has multiple environmental impacts, including greenhouse gas emissions, biodiversity loss, and reduction of other ecosystem services such as carbon storage and water supply (Foley *et al* 2007). Recognizing the environmental, economic

and social importance of these ecosystems and the human potential to change outcomes through improved land management policies, many national and international initiatives to reduce deforestation are underway. Commitments to deforestation-free supply chains have been made by many of the world's largest commodity suppliers (Supply Change/Forest Trends 2015), and recent international agreements include ambitious goals to reduce or end deforestation over the next 5–15 years (UNEP 2010, UN 2015, UN Climate Summit 2014, UNFCCC 2016).



To understand whether we are making progress towards achieving these goals, monitoring systems that deliver timely and accurate information about forest dynamics become critical. Due in part to REDD+⁶ initiatives and guided by simultaneous advances in technology, many governments are developing forest monitoring systems based on the interpretation of imagery from Earth-observing satellites (Romijn *et al* 2015). At the global scale, Hansen *et al* (2013) examined Landsat satellite data to characterize global forest extent, annual loss, and gain from 2000 to 2012. These data and their annual updates through the year 2014, available on the online Global Forest Watch (GFW) platform, provide spatially detailed and timely information on forest

dynamics that is both globally consistent and locally relevant. The GFW platform enhances the practical use of these data by providing solutions to the challenges often associated with big data including visualization, storage, analysis, sharing and querying.

These and other data products derived from satellite imagery have fundamentally changed the way the world's forests are monitored. But as sources of data become larger, more complex and more numerous, the ability to quickly explore and interpret patterns with confidence becomes a critical bottleneck for effectively using these data to inform forest policy and management decisions. The Hansen *et al* (2013) data include billions of individual pixels, and visually evaluating patterns in the data at multiple spatial and temporal scales quickly proves infeasible, even on an online platform optimized for this purpose. For example, figure 1(a) shows forest loss in the Sankuru

⁶ Reducing Emissions from Deforestation and Forest Degradation in developing countries, and the role of conservation, sustainable management of forests, and enhancement of carbon stocks.

Nature Reserve in the Democratic Republic of the Congo (DRC) between 2000 and 2014. At this local scale, spatial patterns of loss are clear enough to identify where to prioritize interventions or target potential future loss. At the provincial (figure 1(b)) or national (figure 1(c)) scale, however, analysis by visual inspection alone no longer provides clear guidance on which areas to target. Furthermore, these visual inspections do not allow an understanding of more subtle, but still important, trends in the acceleration or deceleration of loss.

Several machine learning techniques and land cover change prediction models have been developed to project future deforestation based on where it has occurred historically (Harris *et al* 2008, Fuller *et al* 2011, Rosa *et al* 2013, Aguilar-Amuchastegui *et al* 2014). Despite differences in model structure and complexity, most require at least two land cover maps as well as spatial information about the biophysical and socioeconomic factors that correlate with observed change. These models are conceptually robust, but predictive accuracy is dependent on the availability of spatially-explicit information about drivers of land cover change for model calibration that may not be available.

Spatial pattern analysis, while not inherently predictive, also has the potential to assist in rapid, consistent identification of priority areas for management intervention. For example, statistical methods have been used to identify trends in biodiversity (Myers 1988, Mittermeier 1998, Myers *et al* 2000), pollution (Yong-Hui *et al* 2010, Li *et al* 2014, Ding *et al* 2015), and crime (Grubestic 2006, Chainey *et al* 2008, Xiaoland and Grubestic 2010, Sangamithra *et al* 2012). In the context of forest conservation, spatial statistics can assist in quickly identifying spatiotemporal trends of forest loss without the explicit need for pre-existing information on what underlying factors are driving these trends. It is clear that tropical forest countries must drastically reduce deforestation over the next several years if the goal set in the New York Declaration on Forests to halve natural forest loss by 2020 is to be realized (Zarin *et al* 2016). To succeed, decision makers could benefit from accessible information not just about specific locations of historical forest loss, but also about broader trends in these data that can be detected early enough to influence a new trajectory.

The objective of this study was to develop a scalable methodological approach that allows for rapid evaluation of statistical trends in forest monitoring data. To demonstrate our approach, we use the 14-year annual time series forest loss data for Brazil, Indonesia and the Democratic Republic of the Congo (DRC) from Hansen *et al* (2013). Recent research has pointed out the significant roles these three countries play under various scenarios to reduce deforestation (Zarin *et al* 2016), but did not identify where, geographically, each country should focus policy within its borders to achieve these reductions. An approach that delivers

rapid insight towards answering this question could assist in the creation of more timely and coherent policies on where to focus forest monitoring and enforcement efforts going forward.

2. Methods

2.1. Definitions

The term ‘hot spot’ has been used generically across disciplines to describe a region or value that is higher relative to its surroundings (Lepers *et al* 2005, Aben *et al* 2012, Isobe *et al* 2015). In a forest conservation context, WWF (2015) identified deforestation ‘fronts’ as broad regions of concern based on expert opinion and scenario analyses where available. Here, we define a hot spot as an area that exhibits statistically significant clustering in the spatial pattern of forest loss. Hot spots are locations where observed patterns are not likely the result of random processes or of subjective cartographic design decisions; they represent places where there are underlying spatial processes at work (Getis and Ord 1992). Emerging Hot Spot Analysis extends this definition to incorporate information about the temporal dimension of the data.

In this analysis, forest is defined at the 30 m Landsat pixel scale as areas with the biophysical presence of trees >5 m in height and with tree canopy density greater than 25% in the base year 2000; these areas may take the form of natural forests or tree plantations. Forest loss is defined as a ‘stand replacement disturbance’, meaning the removal or significant reduction of tree cover that can result from a variety of factors globally (Hansen *et al* 2013). Thus ‘forest loss’ as defined here does not always equate to deforestation, i.e. conversion of forest to a new, non-forest land use (FRA 2015, Tropek *et al* 2014). However, deforestation dominates the forest loss dynamic observed within the three study areas chosen for this analysis. We analyzed emerging hot spots of forest loss only, because the forest gain product of Hansen *et al* (2013) is not annual. Furthermore, data from Hansen *et al* (2013) are used here for illustrative purposes; these could be combined with other information on land use or forest type or replaced with other spatiotemporally explicit data where available (e.g. annual deforestation data for the Brazilian Amazon, 1988–2014, INPE 2016) to reflect alternate definitions of forest and forest loss/deforestation.

2.2. Study area

Emerging Hot Spot Analysis could be performed at a global scale using one set of standardized parameters across all geographic domains. However, to be relevant for land use policy, these analyses are likely to be more useful if they relate to dynamics occurring within specific domains of legal responsibility. Brazil, Indonesia and DRC consistently rank within the top ten countries globally for highest absolute forest loss

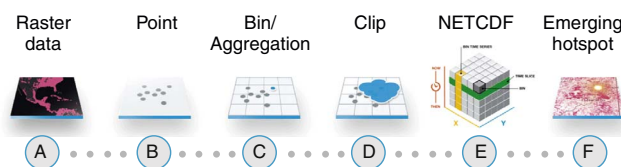


Figure 2. Data processing and analysis workflow for converting raster data (30 m) into an emerging hot spot map. (a) Raster data are stored in a cloud environment; (b) Raster data are converted to points; (c) points are aggregated into space-time bins and point counts are tallied; (d) data are extracted for analysis for an area of interest; (e) data are converted into a netCDF data cube format; and (f) for each bin in the cube, the Getis-Ord G_i^* statistic is calculated, and the Mann-Kendall trend test is calculated for each bin's time series to produce an emerging hot spot map.

(Hansen *et al* 2013), and are home to the largest remaining contiguous expanses of tropical forest (Potapov *et al* 2008). In addition to their conservation value, these countries also represent ideal candidates to test our methodology because each is dominated by a different forest disturbance dynamic, ranging from large tracts of forest cleared in Brazil for industrial agriculture to small patches of forest cleared in DRC to dynamic forest landscapes in Indonesia interspersed with tree plantations managed for industrial oil palm and wood pulp. We focus our analysis at the national scale for these three countries, but also demonstrate the utility of a scalable approach by generating results for a single province as well as a single protected area within DRC.

2.3. Input data and pre-processing

A summary of the processing steps used in our analysis is shown in figure 2. Input data consisted of forest extent and annual forest loss data produced by the University of Maryland and Google (Hansen *et al* 2013 and updates available at globalforestwatch.org). These are delivered as raster data at 30 m spatial resolution. Forest loss in the original data product that occurred outside the forest extent mask applied (>25% tree canopy density) was not considered in the analysis.

Before running statistical analysis, we transformed the data into a netCDF (Network Common Data Form) data cube structure by aggregating forest loss points in each country into space-time 'bins' with a spatial resolution of 2.5 km (figure 2). The value of each bin was assigned as the count (number) of forest loss incidents in the bin for a given year. The netCDF structure stores space as latitude and longitude coordinates and time (i.e. the year the loss was observed) as another dimension. By assigning a count of loss points to each bin for all locations containing forest in the year 2000, the trend in point counts over time could be evaluated. Unlike Brazil and DRC, data for Indonesia were not processed as a country. Rather, we processed Kalimantan and Sumatra (the two Indonesian islands with highest forest cover) independently (i.e. data points on each island did not affect analysis results for the other island) and other island groups were excluded from the analysis. The decision to aggregate data into 2.5 km bins was made after empirically testing bin sizes ranging from 1 km to

50 km; the final selection of 2.5 km preserved a varied frequency distribution of forest loss counts across bins and also preserved a useful interpretation of results that visually tracked known ground features such as lakes, rivers and non-forested areas. Small adjustments to bin size did not significantly impact final results.

Parallel processing and data management were conducted using a combination of several open-source software platforms, including Spark and Hadoop. The Esri ArcGIS Emerging Hot Spot Analysis geoprocessing tool was used for statistical analysis and ESRI software was used for map symbolization. All processing was done in an Amazon Web Services cloud computing environment.

2.4. Statistical analysis

The Emerging Hot Spot Analysis tool evaluates spatiotemporal patterns in forest loss in each country using a combination of two statistical measures: (1) the Getis-Ord G_i^* statistic (Ord and Getis 1995) to identify the location and degree of spatial clustering of forest loss, and (2) the Mann-Kendall trend test (Mann 1945, Kendall and Gibbons 1990) to evaluate temporal trends across the time series.

First, the Getis-Ord G_i^* statistic measures the intensity of clustering of high or low values (i.e. counts of forest loss) in a bin relative to its neighboring bins in the data cube. The sum for a bin and its neighbors is compared proportionally to the sum of all bins. When a bin's sum is different than expected, and that difference is too large to be the result of random chance, a statistically significant Z score is the result. The Getis-Ord G_i^* statistic generates Z scores (standard deviations) and P values (statistical probabilities) for each bin that indicate whether forest loss in a given bin is statistically clustered compared to loss in neighboring bins, as well as loss across the entire analysis domain. A Z score above 1.96 or below -1.96 means that there is a statistically significant hot spot or a statistically significant cold spot of forest loss at a significance level of $P < 0.05$. The larger a bin's Z-score, the more intense the clustering of values (hot spot). Due to the cube structure of the data, neighboring bins exist both in time and in space. The Queens Case Contiguity method was used to define neighborhood size in space and temporal neighbors were defined using one prior time-step interval (1 year).

Second, the Emerging Hot Spot Analysis tool uses the Mann-Kendall statistic (Mann 1945, Kendall and Gibbons 1990) to test whether a statistically significant temporal trend exists through each bin's 14-year time series of Z-scores resulting from the Getis-Ord G_i^* statistic. The most common method for examining the existence of a trend is simple linear regression, a parametric statistical technique that requires independent random samples drawn from normally distributed populations (Taylor and Taylor 1977). However, these requirements may or may not be met in wall-to-wall geospatial data sets (vs. a statistically drawn sample (Schlagel and Newton 1996)). Under these conditions, the use of a non-parametric statistical test is more appropriate. The Mann-Kendall test, an application of Kendall's rank order correlation test to time series data (Bradley 1968), is a nonparametric test for zero slope of the linear regression of time-ordered data versus time (Gilbert 1987).

In this case, each bin with data represented its own independent time series. To evaluate temporal trends for each bin, each time step was compared to the one after it. If the Z-score in the second time step was larger than the first time step, the result was +1 (increasing trend). If the Z-score in the second time step was smaller than the first time step, the result was -1 (decreasing trend). Each pair of time steps was compared over the 14-year series, generating the Mann-Kendall statistic with associated trend Z-score and p-value for each bin. The expected sum is zero, indicating no trend in the values over time. Based on the variance for the values in the bin time series and the number of time periods, the observed sum is compared to the expected sum (zero) to determine if the difference is statistically significant or not ($P < 0.05$). The cluster and trend results from the Getis Ord G_i^* and Mann-Kendall statistics are then used to categorize each bin.

2.5. Output maps

By default, the Emerging Hot Spot Analysis tool categorizes each bin into one of seventeen distinct categories that cover a range of scenarios: one category of non-significance along with eight hot spot and eight cold spot categories, each reflecting a different configuration of spatiotemporal significance (ESRI 2016). However, introducing too much nuance in output maps has the potential to detract attention away from the most relevant results for forest conservation. Thus we chose not to display cold spots and other areas with non-significant trends, and aggregated hot spots into five categories, each representing a different temporal state: new, sporadic, intensifying, persistent, and diminishing (table 1). Resulting hot spot maps for each country were evaluated against other extant data including intact forest landscapes (Potapov *et al* 2008), locations of tree plantations (Petersen *et al* 2016), primary forests (Margono *et al* 2014, Potapov *et al* 2012), and

Table 1. Classification scheme for portraying different categories of statistically significant hot spots, each representing a different temporal state.

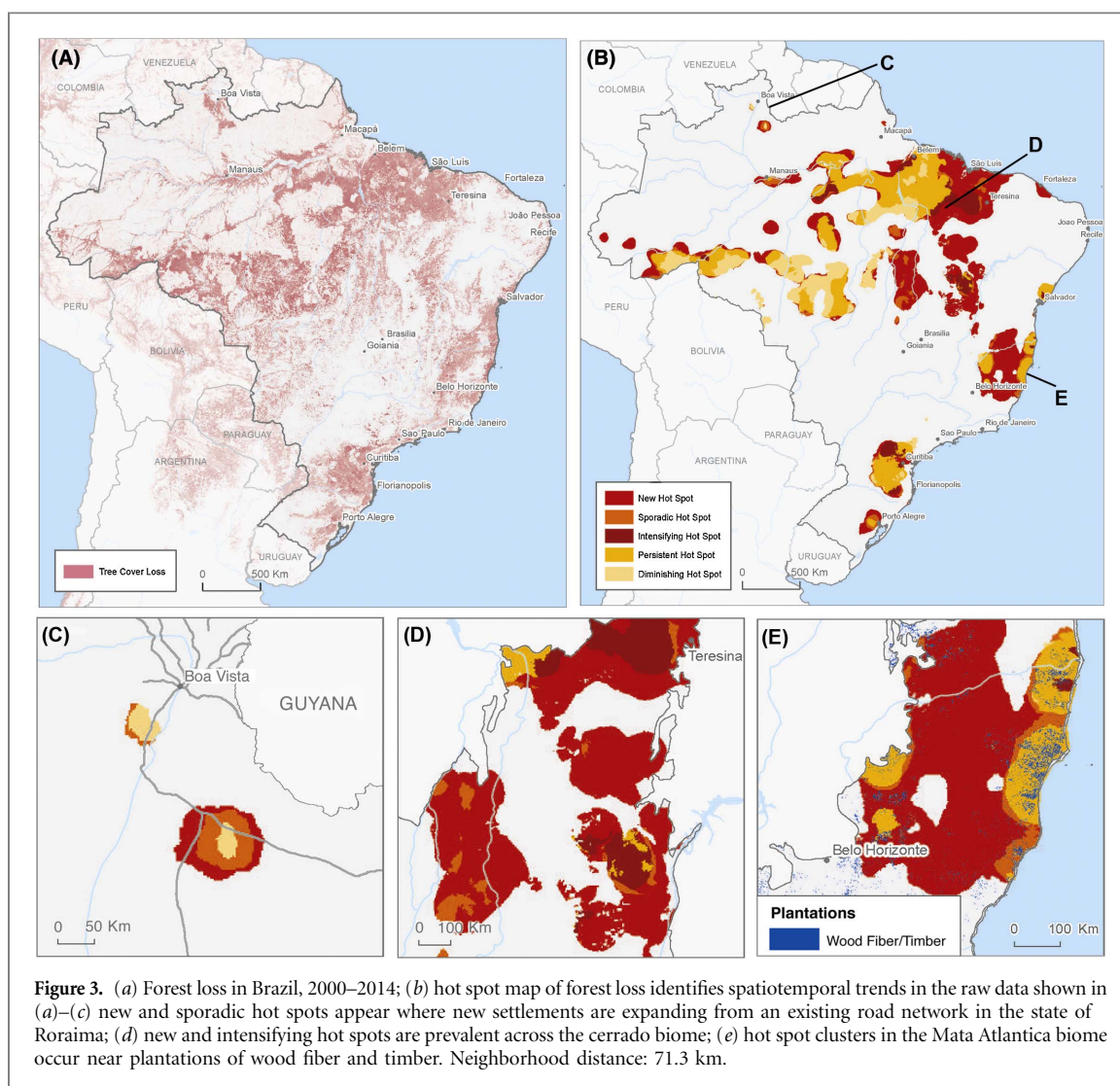
Hot spot category name	Definition
New	A location that is a statistically significant hot spot only for the years 2013 and/or 2014, i.e. the last time steps of the time series.
Sporadic	A location that is an on-again then off-again hot spot. Less than 12 of the 14 years have been statistically significant hot spots.
Intensifying	A location that has been a statistically significant hot spot for more than 12 of the 14 years (>90%), including the final time step (2014). In addition, the intensity of clustering of high counts in each time step is increasing.
Persistent	A location that has been a statistically significant hot spot for more than 12 of the 14 years (>90%), with no discernible trend indicating an increase or decrease in the intensity of clustering over time.
Diminishing	A location that has been a statistically significant hot spot for more than 12 of the 14 years (>90%). In addition, the intensity of clustering of high counts in each time step is decreasing, <u>or</u> the most recent year (2014) is not hot.

protected areas (World Conservation Union and UNEP-World Conservation Monitoring Centre 2007) to better understand the how each hot spot category related to the loss of different types of forest in each country.

3. Results

3.1. Brazil

Between 2000 and 2014, Brazil lost an average of 2.7 Mha yr^{-1} of forest (figure 3(a)), although overall rates of loss declined over this time period. The diminishing hot spot category is prominent within the Amazonian states of Rondônia and Mato Grosso, indicating the deceleration of the clustering of loss there. Meanwhile, the historical 'arc of deforestation' is expanding west through the state of Acre and advancing inwards toward more intact portions of the Amazon (figure 3(b)). A new and significant cluster of forest loss appears in the state of Roraima, where new settlements are expanding from an existing road network (figure 3(c)). Loss has also shifted towards the cerrado biome (Beuchle *et al* 2015), as evidenced by the large areas of new and intensifying hot spots there (figure 3(d)). The overlap of the two predominant hot spot clusters with locations of tree plantations in the southern part of the country in the Mata Atlantica biome likely reflect plantation harvest cycles of the pine and eucalyptus plantations that are prevalent in the states of Parana, Rio Grande do Sul, and Santa Catarina (figure 3(e)).



3.2. Indonesia

Indonesia lost an average of 1.3 Mha yr^{-1} of forest between 2000 and 2014 (figure 4(a)), 39% of which occurred within Indonesia's primary forests (Margono *et al* 2014). Vast areas of new hot spots are present in Sumatra and Kalimantan (figure 4(b)) (Stibig *et al* 2013). Unlike Brazil, both Kalimantan and Sumatra lack diminishing hot spots. Persistent hot spots appear in Riau province in Sumatra and in West Ketawaringin reGENCY in Central Kalimantan, and both show leading fronts of intensified loss clustering (figure 4(b)). In Riau province, an intensifying hot spot pushes through Tesso Nilo National Park and new hot spots overlap with one of the last remaining expanses of intact primary forest in the region (figure 4(c)). Sebangau National Park, the protected peat swamp forest west of Palankaraya in Central Kalimantan, is sandwiched between several new hot spots recently cleared for oil palm plantations (figure 4(d)), and new hot spots are encroaching into the northern end of the park. New hot spots also cover much of West Kalimantan province, where plantation forestry and oil palm expansion are prevalent (figure 4(e)) (Stibig *et al* 2013). Hot spot results for Indonesia are

complicated by the presence of large areas of plantations across both Sumatra and Kalimantan, because harvest cycles within plantations boundaries are observed as forest loss. However, since oil palm is cut and regrown on 25–30 year cycles and the palm oil industry here is relatively nascent, the majority of the identified hot spots likely reflect natural forest loss to establish the plantations, rather than harvest cycles occurring within already existing plantations.

3.3. Democratic republic of the Congo

From 2000 to 2014 DRC lost an average of 0.57 Mha yr^{-1} , and the rate of forest loss between 2011 and 2014 increased by a factor of 2.5 (figure 5(a)). Like Kalimantan and Sumatra in Indonesia, DRC contains vast areas of new hotspots, many of which are situated along the country's road network (figure 5(b)). Unlike Indonesia, however, many more of these new hot spots intersect with intact primary forests. New hot spots also extend within the boundaries of protected areas such as the Sankuru Nature Reserve, the world's largest continuous protected area for great apes (figure 5(c)). New and intensifying hot spots of loss radiate out from the city of Beni in the eastern province of North Kivu,

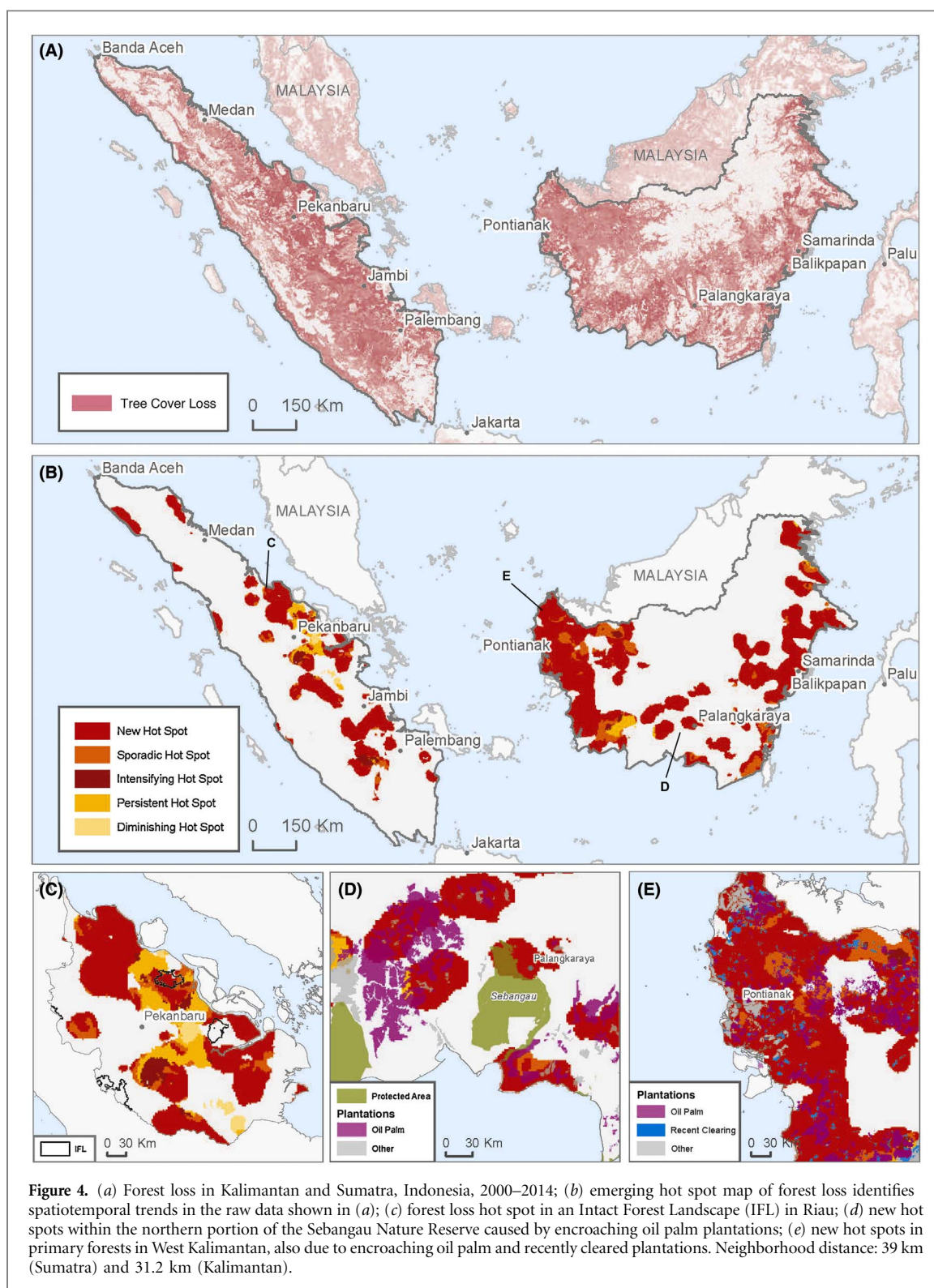


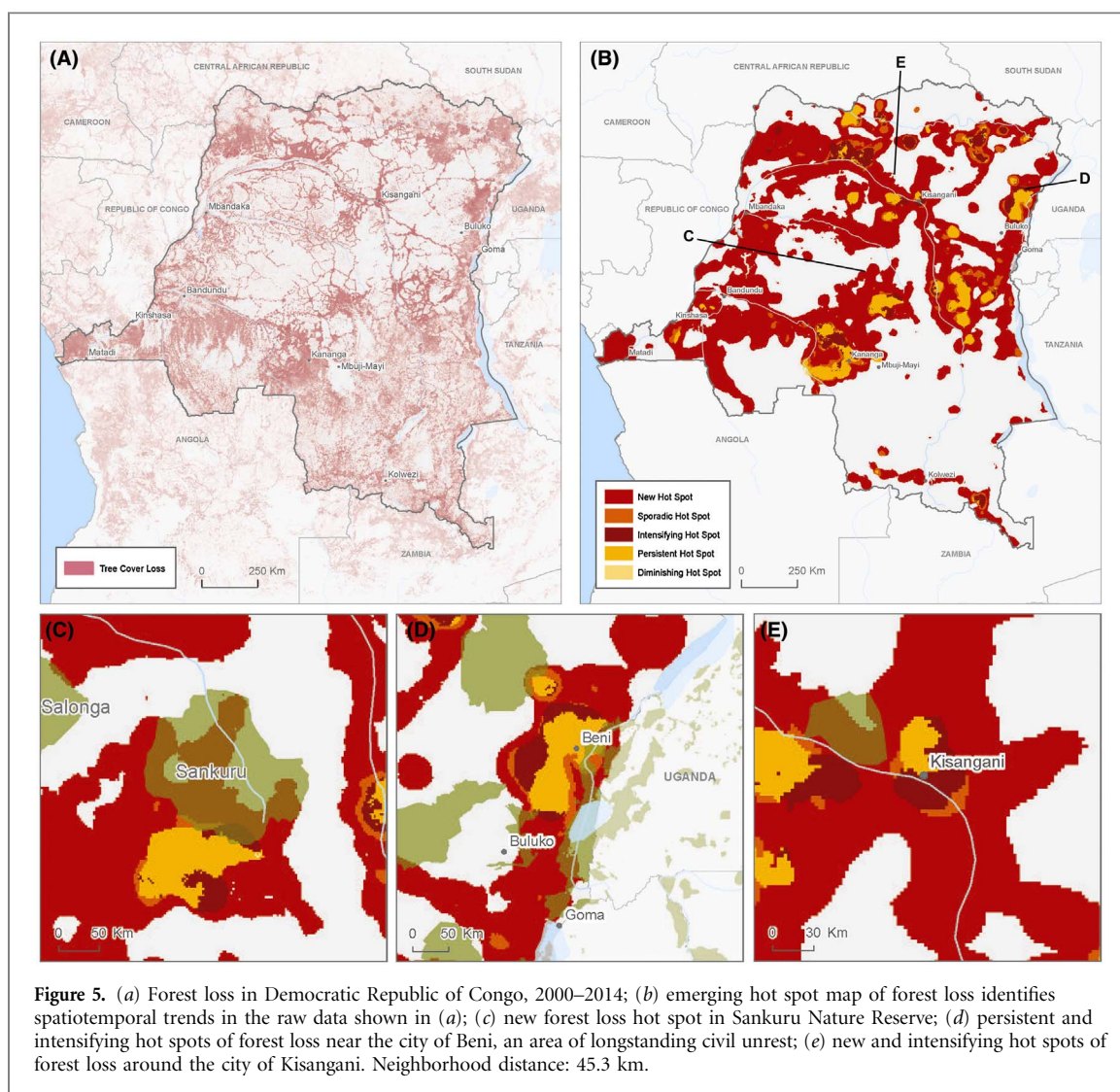
Figure 4. (a) Forest loss in Kalimantan and Sumatra, Indonesia, 2000–2014; (b) emerging hot spot map of forest loss identifies spatiotemporal trends in the raw data shown in (a); (c) forest loss hot spot in an Intact Forest Landscape (IFL) in Riau; (d) new hot spots within the northern portion of the Sebangau Nature Reserve caused by encroaching oil palm plantations; (e) new hot spots in primary forests in West Kalimantan, also due to encroaching oil palm and recently cleared plantations. Neighborhood distance: 39 km (Sumatra) and 31.2 km (Kalimantan).

where longstanding civil unrest has led to a wave of human migration (figure 5(d)). Intensifying hot spots are also apparent near city centers such as Kisangani, where mining as well as fuelwood collection and shifting agriculture practiced by local communities is common (figure 5(e)). The presence of intensifying hot spots in these locations suggests the acceleration of forest loss clustering there, potentially as a result of forest conversion to new agricultural land as part of an expanding rural complex to accommodate DRC's

rapidly growing population (Mayaux *et al* 2013, Molinario *et al* 2015).

4. Discussion

Here we demonstrate the usefulness of Emerging Hot Spot Analysis as an approach for evaluating statistically significant clusters of forest loss. Unlike traditional predictive modeling approaches, our method and workflow can be implemented quickly and requires no



information on the underlying drivers of loss, yet results can provide insight about potential future trajectories and locations of forest change.

We apply this method at the national scale for three tropical forested countries over 14 years, but the approach is scalable to any spatial or temporal domain with more than ~ 10 time steps. For very local applications, this type of analysis may not offer much advantage over viewing the data in its native resolution (in this case, 30 m). Emerging Hot Spot Analysis becomes much more valuable for evaluating phenomena occurring over broader scales than can be directly observed, and where limited time and resources may necessitate the prioritization of monitoring and enforcement efforts to target specific fronts of new or accelerated loss. Decision makers tasked with vetting proposed forest protection policies or commodity supply chain sourcing decisions through multiple stakeholder groups may benefit from the support of maps that are easy to interpret, yet grounded in robust statistical analysis. National and subnational governments seeking to obtain and receive results-based payments for REDD+ activities under the UNFCCC may also find the emerging hot

spot framework useful for informing their national strategies and action plans.

The fixed scale of an emerging hot spot analysis is a critical concept to understand for proper interpretation of results, because emerging hot spot maps are dependent upon the extent of the analysis domain under consideration. For example, figure 6 shows how emerging hot spot maps differ slightly depending on the three different analysis domains defined in figure 1: a single protected area, a province, and all forests of DRC. Output maps differ in part because results for each bin are calculated relative to different 'global means', representing the average loss count across the entire analysis domain.

Results are also sensitive to neighborhood size, or the distance over which each point is compared to all others. As neighborhood size increases, hot spots will become larger and fewer; smaller neighborhood sizes capture more localized trends. There is inevitably an element of subjectivity in choosing an appropriate value for neighborhood distance; Atkinson and Unwin (2002) state that 'subjective judgement based on a range of density surfaces is as good a method as any' for determining an appropriate neighborhood size. In

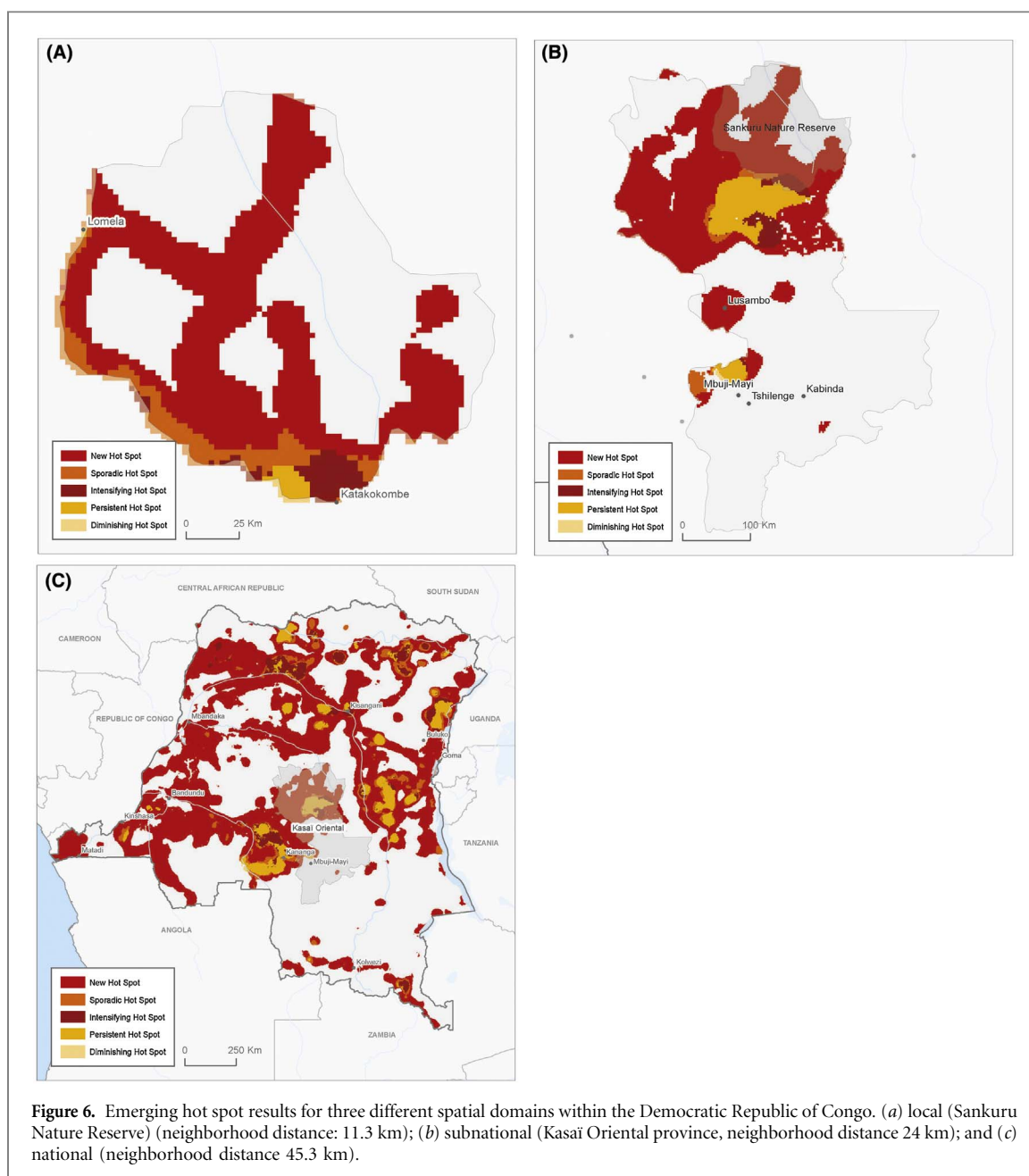
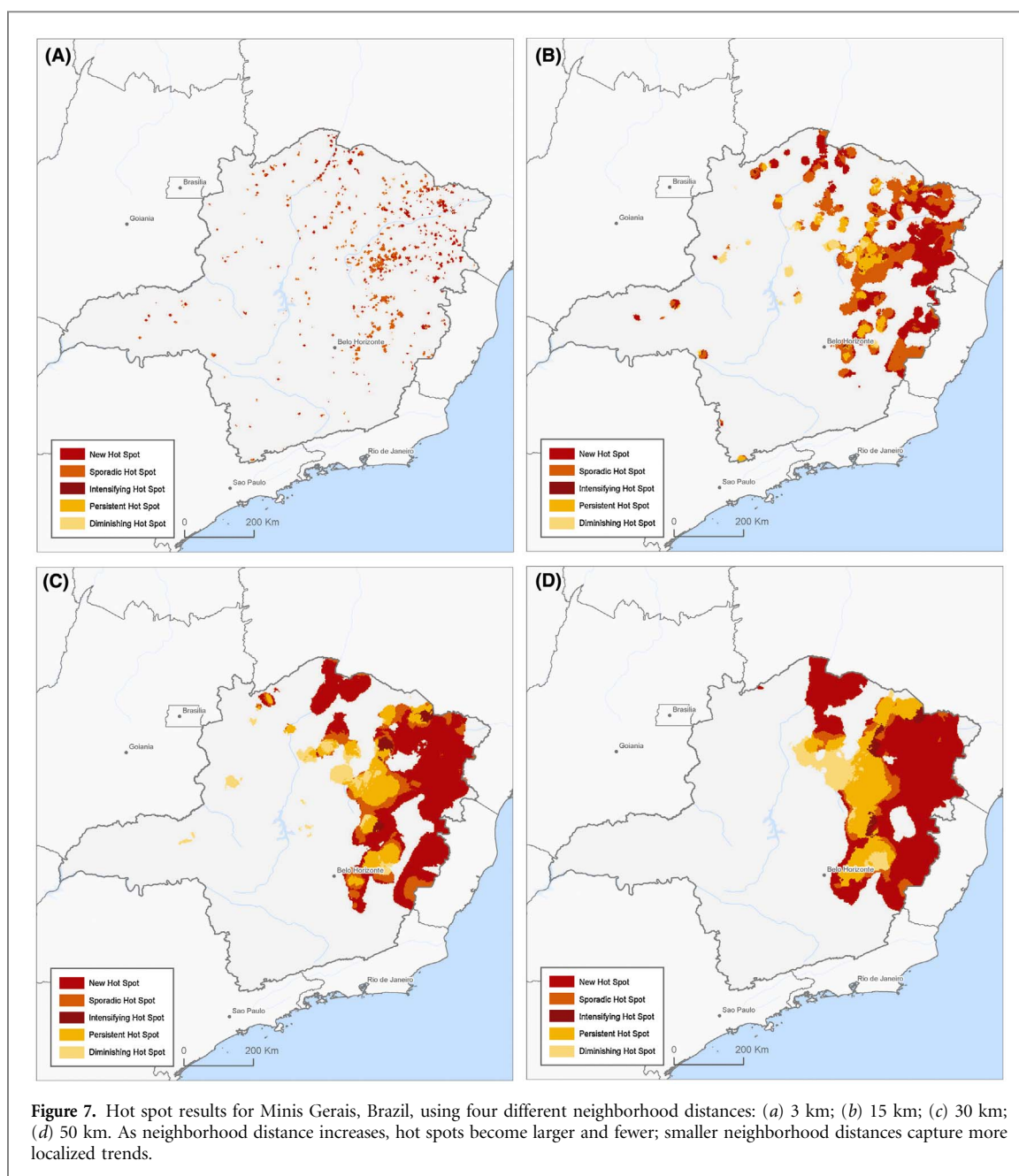


Figure 6. Emerging hot spot results for three different spatial domains within the Democratic Republic of Congo. (a) local (Sankuru Nature Reserve) (neighborhood distance: 11.3 km); (b) subnational (Kasaï Oriental province, neighborhood distance 24 km); and (c) national (neighborhood distance 45.3 km).

practice, the scale of the analysis should influence the neighborhood size, and specific neighborhood distances used in this analysis are provided in figures 3–6. A range of neighborhood distances, and the impact on resulting maps, is also shown in figure 7. Despite the variations that emerge from the use of different neighborhood distances, locations of emerging hot spots are broadly consistent across maps. Using an intermediate neighborhood distance relative to the scale of analysis, as we do for generating results in figures 3–6, highlights specific local hot spots while also allowing other more general patterns to emerge.

All spatial clustering approaches, regardless of their theoretical underpinning, statistical foundation, or mathematical specification, have limitations in accuracy, sensitivity, and the computational effort required for identifying clusters. As a result, a major challenge in practice is determining which technique(s)

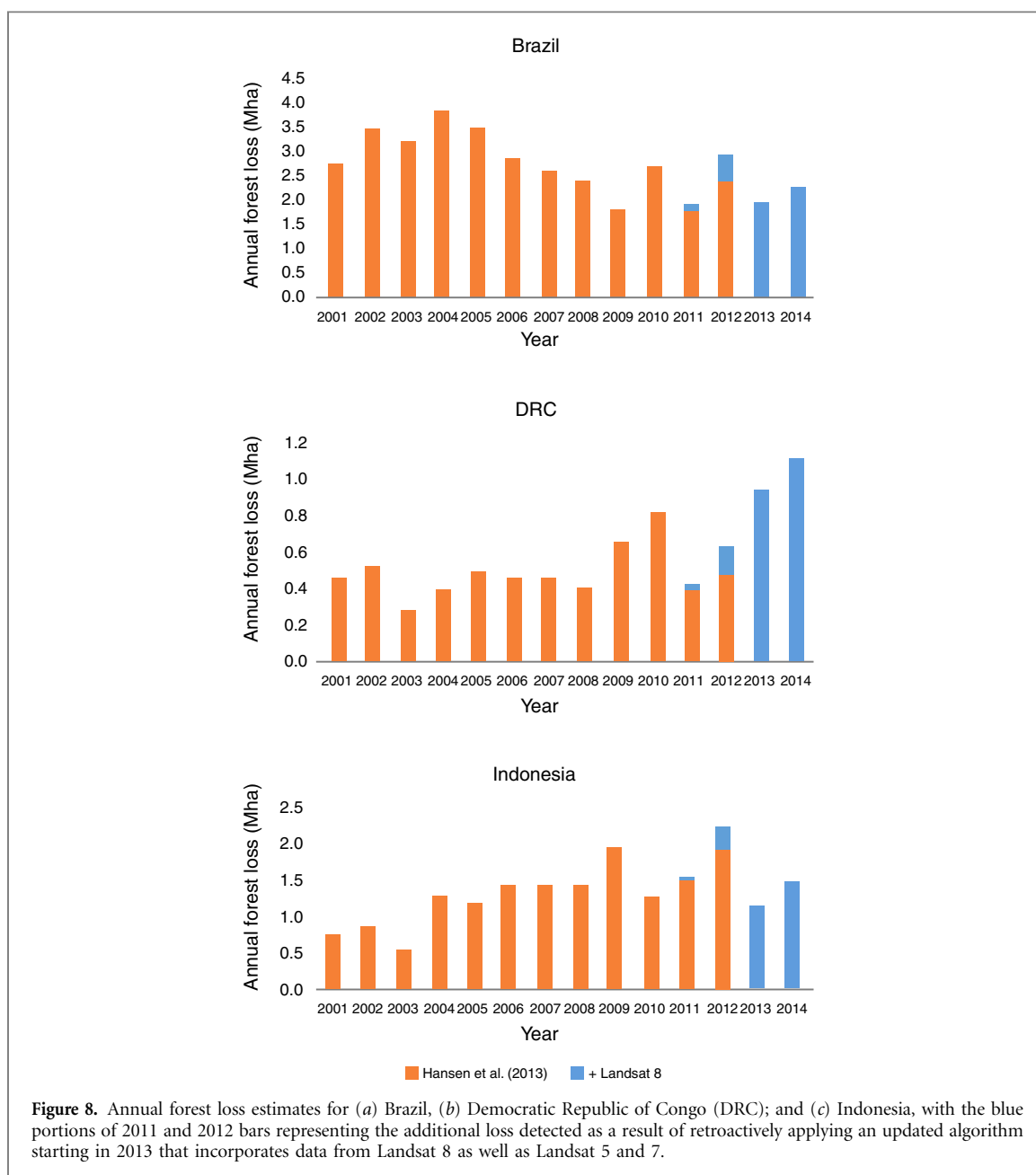
will provide the most meaningful insights for a particular issue or context (Grubestic *et al* 2014). Contextual data layers are therefore extremely important for interpreting results, as is knowledge about the accuracy of the underlying data product used to generate statistical output. Across the tropical biome, user’s and producer’s accuracies of the forest loss data used for this analysis are 87.0 and 83.1%, respectively (Hansen *et al* 2013). Another more detailed accuracy assessment by Tyukavina *et al* (2015) shows that the original Hansen *et al* (2013) data product (version 1.0) underestimates loss in Brazil by 9%, overestimates loss in Indonesia by 9%, and underestimates loss in DRC by 65%. Lower accuracy in DRC was due to missed loss in landscapes dominated by smallholder rotation agriculture. Since original publication, the forest loss data of Hansen *et al* (2013) have been updated twice, and version 1.2 now includes loss up to the year 2014. The



latest version of the global dataset includes forest loss data re-processed from year 2011 using a new change detection model that is more sensitive to forest loss and includes gap-free Landsat 8 data since 2013. These changes resulted in improved detection of loss in DRC and other areas where smallholder agricultural clearing, selective logging, and short cycle plantation clearing are prevalent. However, the years preceding 2011 have not been reprocessed in this manner. Thus the higher sensitivity of the new change detection model and the use of Landsat 8 data since 2013 is likely driving some, but not all, of the prevalence of 'new' hot spots in our results, particularly in DRC. The additional loss detected in 2011–2012 for each country as a result of retroactively applying the updated change detection algorithm is shown in figure 8. While this temporal inconsistency is a clear limitation of our analysis results, future planned research includes the

full re-processing of the time series back to the year 2000.

The framework outlined here demonstrates the added analytical value offered by forest monitoring data with high temporal cadence, or resolution. While we use a widely cited data product to demonstrate our approach (Hansen *et al* 2013) in three countries with relatively well studied forest loss narratives, any other data with a space and a time component can be evaluated using the same method. Because the Hansen *et al* (2013) forest loss product is binary (i.e. values represent the presence or absence of loss), counts of forest loss within each bin were used as the aggregation method. Depending on the specific research question, similar analysis could be run on other binary data products, such as active fire points (NASA FIRMS 2016) or weekly forest disturbance alerts (Hansen *et al* 2016). Data could also be aggregated by value rather



than by count for continuous variables; for example, for applications targeting locations for interventions to reduce emissions from deforestation (REDD+), rather than the area of deforestation, bin values could reflect the amount of biomass lost within a bin in a given year. In this example, point locations and years would be the same as for the forest loss data of Hansen *et al* (2013), but resulting hot spot maps would differ based on the aggregation method, especially in areas where there is substantial variation in forest biomass, such as in the Amazon vs. cerrado biomes of Brazil (figure 9) (Baccini *et al* 2012, Ometto *et al* 2013, Achard *et al* 2014). In the emerging hot spot map of biomass loss, the quantity of biomass lost within a bin drives the analysis, rather than the presence or absence of loss.

The maximum domain over which these types of analyses can be run was formerly defined by the processing power of the machine used to run the analyses. Recent and rapid advances in cloud

computing and parallel processing have removed this barrier, such that complex data analysis over both small and large regions is possible and sensitivity of results to different parameters can be assessed. We have demonstrated an approach for quickly and dynamically assessing emerging clusters of forest loss, and the analysis can be re-run for new temporal windows as additional time steps are added. Moving forward, analysis of forest monitoring data using the presented approach could help corporate sustainability officers, environmental groups, resource managers, and governments gain rapid insights on where to focus conservation and management interventions for custom areas and time periods of interest.

5. Conclusions

The acquisition of satellite data with decreased latency times is providing massive streams of near real-time

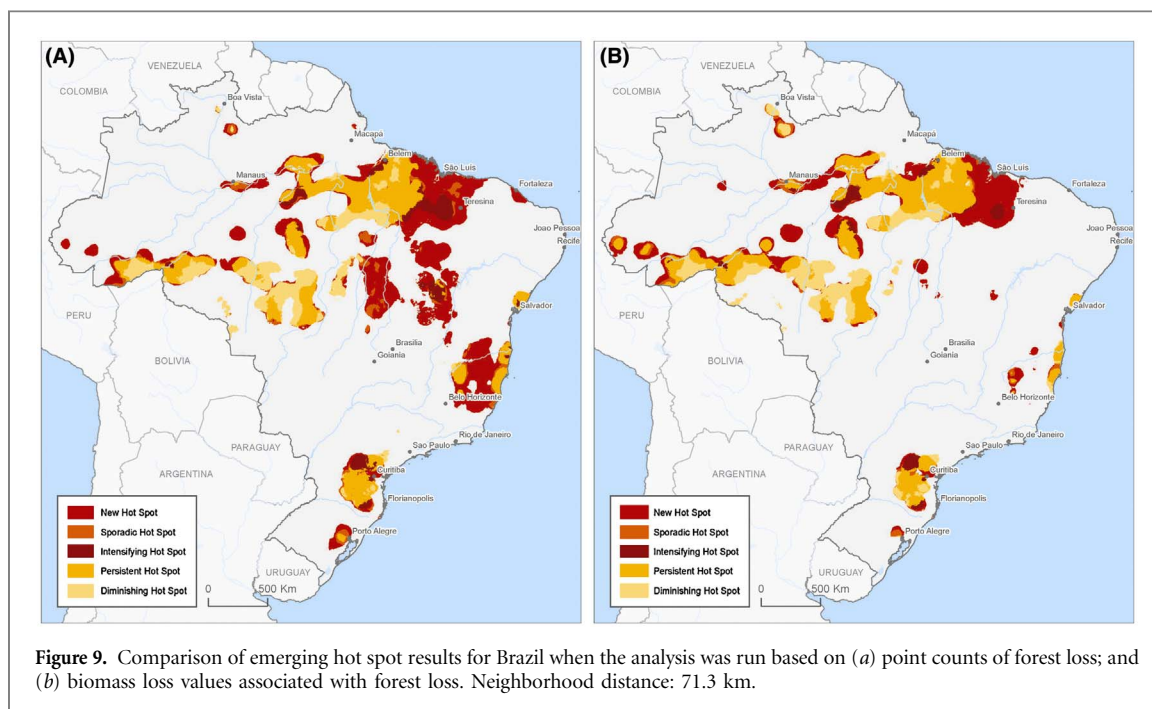


Figure 9. Comparison of emerging hot spot results for Brazil when the analysis was run based on (a) point counts of forest loss; and (b) biomass loss values associated with forest loss. Neighborhood distance: 71.3 km.

data to the remote sensing research community. The key to applying these new data streams in a forest conservation context will be to process and analyze them quickly, turning raw data into actionable insights that facilitate timely and well informed management decisions. Continued innovation in both software and cloud computing will enable the approach outlined here to be applied in novel ways to allow greater understanding of, and better management of, complex environmental systems.

Acknowledgments

This study was made possible in part by the UK Department for International Development (DFID Purchase Order 40078074), the US Agency for International Development (AID-OAA-A-13-00045) and the Norwegian Ministry of Foreign Affairs (QZA-13/0313) in support of the Global Forest Watch (GFW) Partnership convened by the World Resources Institute (globalforestwatch.org).

References

- Aben J, Adriaensens F, Thijs K W, Pellikka P, Siljander M, Lens L, Matthyens E, Altwegg R and Bessa-Gomes C 2012 Effects of matrix composition and configuration on forest bird movements in a fragmented Afrotropical biodiversity hot spot *Anim. Conserv.* **15** 658–68
- Achard F *et al* 2014 Determination of tropical deforestation rates and related carbon losses from 1990 to 2010 *Glob. Change Biol.* **20** 2540–54
- Aguilar-Amuchastegui N, Riveros J C and Forrest J L 2014 Identifying areas of deforestation risk for REDD+ using a species modeling tool *Carbon Balance Manage.* **9** 1–10
- Atkinson P J and Unwin D J 2002 Density and local attribute estimation of an infectious disease using MapInfo *Comput. Geosci.* **28** 1095–106
- Baccini A *et al* 2012 Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps *Nat. Clim. Change* **2** 182–5
- Beuchle R, Grecchi R C, Shimabukuro Y E, Seliger R, Eva H D, Sano E and Achard F 2015 Land cover changes in the Brazilian Cerrado and Caatinga biomes from 1990 to 2010 based on a systematic remote sensing sampling approach *Appl. Geogr.* **58** 116–27
- Bradley J V 1968 *Distribution-free Statistical Tests* (Englewood Cliffs, NJ: Prentice-Hall)
- Chainey S, Tompson L and Uhlig S 2008 The utility of hotspot mapping for predicting spatial patterns of crime *Security Journal* **21** 4–28
- Ding L, Chen K, Liu T, Cheng S and Wang X 2015 Spatial-temporal hotspot pattern analysis of provincial environmental pollution incidents and related regional sustainable management in china in the period 1995–2012 *Sustainability* **7** 14385–407
- ESRI 2016 ‘How Emerging Hot Spot Analysis Works’ (<http://desktop.arcgis.com/en/arcmap/10.3/tools/space-time-pattern-mining-toolbox/learnmoreemerging.htm>) (Accessed: 12 March 2016)
- Foley J A *et al* 2007 Global consequences of land use *Science* **309** 570
- FRA 2015 Terms and definitions. Forest resources assessment working paper 180 (Rome: FAO)
- Fuller D O, Hardiono M and Meijaard E 2011 Deforestation projections for carbon-rich peat swamp forests in Central Kalimantan *Indonesia Environmental Management* **48** 436–47
- Getis A and Ord K 1992 The analysis of spatial association by use of distance statistics *Geogr. Anal.* **24** 189–206
- Gilbert R O 1987 *Statistical Methods for Environmental Pollution Monitoring* (New York: Wiley) ch 16
- Grubestic T 2006 On the application of fuzzy clustering for crime hot spot detection *J. Quant. Criminol.* **22** 77–105
- Grubestic T H, Wei R and Murray A T 2014 Spatial clustering overview and comparison: accuracy, sensitivity and computational expense *Ann. Assoc. Am. Geogr.* **104** 1134–55
- Hansen M C, Krylov A, Tyukavina A, Potapov P V, Turubanova S, Zutta B, Ifo S, Margono B, Stolle F and Moore R 2016 Humid tropical forest disturbance alerts using Landsat data *Environ. Res. Lett.* **11** 034008
- Hansen M C *et al* 2013 High-resolution global maps of 21st-century forest cover change *Science* **342** 850–4

- Harris N L, Petrova S, Stolle F and Brown S 2008 Identifying optimal areas for REDD intervention: East Kalimantan as a case study *Environ. Res. Lett.* **3** 035006
- INPE 2016 Projeto PRODES: Monitoramento da floresta Amazônica Brasileira por satélite (www.obt.inpe.br/prodes/index.php) (Accessed: 12 March 2016)
- Isobe A, Uchida K, Tokai T and Iwasaki S 2015 East Asian seas: a hot spot of pelagic microplastics *Mar. Pollut. Bull.* **101** 618–23
- Kendall M G and Gibbons J D 1990 *Rank Correlation Methods* (London: Griffin)
- Lepers E, Lambin E, Janetos A, DeFries R, Achard F, Ramankutty N and Scholes R 2005 A synthesis of information on rapid land-cover change for the period 1981–2000 *BioScience* **55** 115–24
- Li W, Xu B, Song Q, Liu X, Xu J and Brookes P 2014 The identification of ‘hotspots’ of heavy metal pollution in soil-rice systems at a regional scale in eastern China *Sci. Total Environ.* **472** 407–20
- Mann H B 1945 Nonparametric tests against trend *Econometrica* **13** 245–59
- Margono B A, Potapov P, Turubanova S, Stolle F and Hansen M 2014 Primary forest cover loss in Indonesia over 2000–2012 *Nat. Clim. Change* **4** 730–5
- Mayaux P *et al* 2013 *Phil. Trans. R. Soc. B* **368** 1–10
- Mittermeier R A, Myers N, Thomsen J B, da Fonseca G A B and Silvie Olivieri 1998 Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities *Conserv. Biol.* **12** 516–20
- Molinario G, Hansen M C and Potapov P V 2015 Forest cover dynamics of shifting cultivation in the Democratic Republic of Congo: a remote sensing-based assessment for 2000–2010 *Environ. Res. Lett.* **10** 1–15
- Myers N 1988 Threatened biotas: ‘hot spots’ in tropical forests *Environmentalist* **8** 187
- Myers N, Mittermeier R A, Mittermeier C G, da Fonseca G A B and Kent J 2000 Biodiversity hotspots for conservation priorities *Nature* **403** 853–8
- NASA FIRMS Active Fires Accessed through Global Forest Watch on May 5th 2016 (www.globalforestwatch.org)
- Ord K and Getis A 1995 Local spatial autocorrelation statistics: distributional issues and an application *Geogr. Anal.* **27** 286–306
- Ometto J P, Aguiar A P, Assis T, Soler L, Valle P, Tejada G, Lapola D and Meir P 2013 Amazon forest biomass density maps: tackling the uncertainty in carbon emission estimates *Clim. Change* **124** 545–60
- Petersen R, Aksenov D, Goldman E, Sargent S, Harris N, Manisha A, Esipova E, Shevade V and Loboda T 2016 *Mapping Tree Plantations with Multispectral Imagery: Preliminary Results for Seven Tropical Countries* (Washington, DC: World Resources Institute) (www.wri.org/publication/mapping-tree-plantations) (Accessed: 22 June 2016)
- Potapov P *et al* 2008 Mapping the world’s intact forest landscapes by remote sensing *Ecology and Society* **13** 1–16
- Potapov P, Turubanova S, Zhuravleva I, Hansen M, Yaroshenko A and Manisha A 2012 Forest cover change within the Russian European North after the breakdown of the Soviet Union 1990–2005 *Int. J. Forestry Research* **2012** 1–11
- Ramankutty N, Evan A T, Monfreda C and Foley J A 2008 Farming the planet: 1. geographic distribution of global agricultural lands in the year 2000 *Glob. Biogeochem. Cycles* **22** 30–48
- Romijn E, Lantican C B, Herold M, Lindquist E, Ochieng R, Wijaya A, Murdiyarso D and Verchot L 2015 Assessing change in national forest monitoring capacities of 99 tropical countries *For. Ecol. Manage.* **352** 109–23
- Rosa I M D, Purves D, Souza C and Ewers R 2013 Predictive modeling of contagious deforestation in the Brazilian Amazon *PLoS ONE* **8** 1–14
- Sangamithra V, Kalaikumaran T and Karthik S 2012 Data mining techniques for detecting the crime hotspot by using GIS *International Journal of Advanced Research in Computer Engineering and Technology* **10** 228–31
- Schlagel J D and Newton C M 1996 A GIS-based statistical method to analyze spatial change *American Society for Photogrammetry and Remote Sensing* **62** 839–44
- SEEG 2014 Sistema de Estimativa de Emissão de Gases de Efeito Estufa (<http://seeg.eco.br/en/2014/12/06/seeg-2014-engl/>) (Accessed: 1 July 2016)
- Supply Change/Forest Trends 2015 Firm Commitments: Tracking Company Endorsers of the New York Declaration on Forests (www.forest-trends.org/documents/files/doc_4979.pdf) (Accessed: 6 July 2016)
- Stibig H J, Achard F, Carboni S, Rasi R and Miettinen J 2013 Change in tropical forest cover of Southeast Asia from 1990 to 2010 *Biogeosciences* **10** 12625
- Taylor L R and Taylor R A J 1977 Aggregation, migration and population dynamics *Nature* **265** 415–21
- Tropek R, Sedláček O, Beck J, Keil P, Musilová Z, Šimová I and Storch D 2014 Comment on ‘High-resolution global maps of 21st-century forest cover change’ *Science* **344** 981–1
- Tyukavina A, Baccini A, Hansen M C, Potapov P V, Stehman S V, Houghton R A, Krylov A M, Turubanova S and Goetz S J 2015 Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012 *Environ. Res. Lett.* **10** 1
- UNEP 2010 Strategic Plan for Biodiversity 2011–2020 and the Aichi Biodiversity Targets. COP 10 Decision X/2, UNEP/CBD/COP/DEC/X/2 (www.cbd.int/doc/decisions/cop-10/cop-10-dec-02-en.pdf) (Accessed: 5 July 2016)
- United Nations 2015 *Transforming our World: The 2030 Agenda for Sustainable Development* (New York: United Nations Publications) (www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E) (Accessed: 22 June 2016)
- UNFCCC 2016 Report of the Conference of the Parties on its twenty-first session, held in Paris from 30 November to 13 December 2015 (<http://unfccc.int/resource/docs/2015/cop21/eng/10a01.pdf>)
- UN Climate Summit 2014 Forests: Action Statements and Action Plans (www.un.org/climatechange/summit/wp-content/uploads/sites/2/2014/07/New-York-Declaration-on-Forest-E2%80%93Action-Statement-and-Action-Plan.pdf) (Accessed: 5 July 2016)
- World Conservation Union and UNEP-World Conservation Monitoring Centre 2016 *World Database on Protected Areas* User Manual version 1.4 (Cambridge: WCMC) (http://wdpa.s3.amazonaws.com/WDPManual/English/WDPManual_1_4_EN_FINAL.pdf) (Accessed: 31 January 2017)
- WWF 2015 Saving forests at risk *Living Forests Report* (World Wildlife Fund for Nature) Gland
- Xiaoland W and Grubestic T H 2010 Identifying irregularly shaped crime hot-spots using a multiobjective evolutionary algorithm *J. Geogr. Syst.* **12** 409–33
- Yong-Hui Y, Feng Z, Huai-Cheng G, Hu S, Hui L, Xu D and Cheng-Jie H 2010 Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods *Environ. Monit. Assess.* **170** 407–16
- Zarin D J *et al* 2016 Can carbon emissions from tropical deforestation drop by 50% in 5 years? *Glob. Change Biol.* **22** 1336–47