This is a repository copy of *A machine learning approach to map tropical selective logging*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/139563/

Version: Published Version

**Article:**
Hethcoat, M., Edwards, D., Carreiras, J.M.B. orcid.org/0000-0003-2737-9420 et al. (3 more authors) (2019) A machine learning approach to map tropical selective logging. Remote Sensing of Environment, 221. pp. 569-582. ISSN 0034-4257

https://doi.org/10.1016/j.rse.2018.11.044

---

**Reuse**
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
A machine learning approach to map tropical selective logging

M.G. Hethcoat, D.P. Edwards, J.M.B. Carreiras, R.G. Bryant, F.M. França, S. Quegan

**ARTICLE INFO**

**Keywords:**
- Brazil
- Conservation
- Degradation
- Landsat
- Random Forest
- Selective logging
- Surface reflectance
- Texture measures
- Tropical forests

**ABSTRACT**

Hundreds of millions of hectares of tropical forest have been selectively logged, either legally or illegally. Methods for detecting and monitoring tropical selective logging using satellite data are at an early stage, with current methods only able to detect more intensive timber harvest (> 20 m² ha⁻¹). The spatial resolution of widely available datasets, like Landsat, have previously been considered too coarse to measure the subtle changes in forests associated with less intensive selective logging, yet most present-day logging is at low intensity. We utilized a detailed selective logging dataset from over 11,000 ha of forest in Rondônia, southern Brazilian Amazon, to develop a Random Forest machine-learning algorithm for detecting low-intensity selective logging (< 15 m² ha⁻¹). We show that Landsat imagery acquired before the cessation of logging activities (i.e., the final cloud-free image of the dry season during logging) was better at detecting selective logging than imagery acquired at the start of the following dry season (i.e., the first cloud-free image of the next dry season). Within our study area the detection rate of logged pixels was approximately 90% (with roughly 20% commission and 8% omission error rates) and approximately 40% of the area inside low-intensity selective logging tracts were labelled as logged. Application of the algorithm to 6152 ha of selectively logged forest at a second site in Pará, northeast Brazilian Amazon, resulted in the detection of 2316 ha (38%) of selective logging (with 20% commission and 7% omission error rates). This suggests that our method can detect low-intensity selective logging across large areas of the Amazon. It is thus an important step forward in developing systems for detecting selective logging pan-tropically with freely available data sets, and has key implications for monitoring logging and implementing carbon-based payments for ecosystem service schemes.

1. Introduction

Earth’s tropical forests are being rapidly lost and degraded by agricultural expansion and commercial logging operations, with population growth projected to further increase pressures on forests globally (Asner et al., 2005; DeFries et al., 2010). The ability to monitor forest disturbances is an important component in sustainable forest management, understanding the global carbon budget, and implementing climate policy initiatives, such as the United Nation’s (UN) Reducing Emissions from Deforestation and Forest Degradation (REDD +) programme, which seeks to mitigate climate change and biodiversity losses through improved forest management practices (GOFC-GOLD, 2016). The UN anticipates that payments to nations under REDD + initiatives, which compensate countries for conserving forests (and sequestering carbon), could reach $30 billion annually (Phelps et al., 2010, UN-REDD Programme, http://www.un-redd.org).

Remote sensing is considered the most accurate and cost-effective way to systematically monitor forests at broad spatial scales (Achard et al., 2007; Herold and Johns, 2007; Shimabukuro et al., 2014). Large-scale monitoring of deforestation has significantly improved in recent years, and forest losses can be identified with accuracies > 90% using freely available satellite data (Hansen et al., 2013). In addition, near real-time deforestation tracking and alert systems are now possible with systems like DETER (Shimabukuro et al., 2014), FORMA (Hammer et al., 2014; Hansen et al., 2013), and Global Forest Watch (Hansen et al., 2016). In contrast, methods for detecting and monitoring forest degradation are less developed. Forest degradation is an ambiguous term, with over 50 different definitions and no internationally...
established description (Ghazoul et al., 2015; Simula, 2009). This makes generalizing its impacts difficult, in part because degradation can include forests subject to varying intensities of selective logging, fire, artisanal gold mining, fuelwood extraction, etc., which has hampered the development of coordinated international forest policies to track and monitor forest degradation (Ghazoul et al., 2015; Sasaki and Putz, 2009).

Here we focus on detecting a key driver of forest degradation globally, commercial logging operations. In contrast to forest clearance (i.e. deforestation), selective logging represents a more diffuse disturbance wherein only a subset of trees (typically the most economically valuable) are harvested (Fisher et al., 2014; Putz et al., 2001). The resulting forest maintains some degree of its original composition (e.g. canopy cover, biodiversity measures, carbon content, etc.) but is punctured by treefall gaps and logging roads and consequently lies on a continuum between primary forest and complete deforestation (Ghazoul et al., 2015; Thompson et al., 2013). The intensity of selective logging operations can vary in two main ways: (1) the volume of wood harvested typically ranges up about 50 m$^3$ ha$^{-1}$, as high as 150 m$^3$ ha$^{-1}$ in Asia (Buruivalova et al., 2014; Putz et al., 2001) and (2) the degree to which reduced-impact logging is practiced, in which damage to the remaining forest is minimized by careful planning of road networks, skid trails, and directional felling of trees to limit additional tree or canopy damage (Putz and Pinard, 1993). We acknowledge wood biomass can vary substantially across forest types globally and may not, itself, be a perfect indicator of logging intensity. However, in this manuscript we define logging intensity in terms of wood volume extracted to be consistent with legal restrictions outlined in the Brazilian forest code.

Selective logging activities are often the first anthropogenic disturbance to affect primary tropical forests (Asner et al., 2009b; Nepstad et al., 1999) and are thought to be a major source of carbon emissions from degradation (Hosonuma et al., 2012; Pearson et al., 2017). Moreover, road networks associated with logging are often precursors to additional land-use changes (such as agricultural conversion or development of human settlements) and facilitate further degradation (e.g. increased susceptibility to fires or illegal logging) and forest losses (Alamgir et al., 2017; Kumar et al., 2014; Matricardi et al., 2010). Estimates suggest over 400 million ha of tropical forest, an area the size of the European Union, are earmarked in the tropical timber estate to be logged (Blaser et al., 2011). However, the extent of forest subjected to selective logging across the tropics has yet to be estimated (Asner et al., 2005).

Several authors have tried to address the challenges of using satellite data to estimate forest disturbances from selective logging in the tropics (Asner et al., 2002, 2004a, 2005; Matricardi et al., 2007, 2010; Shimabukuro et al., 2014; Souza and Barreto, 2000; Souza et al., 2005). The majority of approaches employ classification of fractional images derived from spectral unmixing of Landsat scenes. Despite these advancements, Landsat imagery has been considered too coarse to detect large canopy gaps and an abundance of spectrally distinct features, like log landing decks or large road networks, their respective abilities to detect lower logging intensities are unknown. Therefore, using Landsat data to map and quantify selective logging at lower logging intensities (< 20 m$^3$ ha$^{-1}$) remains a major challenge, and the amount of forest disturbance overlooked using currently available techniques is unknown. Yet, growing concerns over the impacts of selective logging on carbon and biodiversity (Bicknell et al., 2014; Edwards et al., 2014; França et al., 2017; Martin et al., 2015; Putz et al., 2008) has led to increased use of improved forest management practices, such as reduced-impact logging (Putz and Pinard, 1993). Consequently, the extent of tropical forests being logged at lower intensities and with reduced-impact is almost certainly expanding. In addition, there is an ever-increasing need to detect and account for the estimated 50–90% of tropical timber on the international market harvested illegally at very low intensities (Branclalion et al., 2018; Kleinschmit et al., 2016). Therefore methods to detect subtle forest disturbances from satellite systems with regular global coverage are urgently needed, both to establish reference levels from historical data (e.g. the vast amount of freely available Landsat archives) and to obtain maximum benefit from current and future systems, such as Landsat 8, 9 and Sentinel-2 (Drusch et al., 2012; Roy et al., 2014).

The primary objective of this study was to develop a new method for detecting selective logging in moist tropical forest with Landsat data. It focuses on reduced-impact selective logging of intensity < 15 m$^3$ ha$^{-1}$ (1–2 trees ha$^{-1}$), much lower than is typically reported in studies that use remote sensing data to estimate selective logging (Asner et al., 2004a, 2005; Souza and Roberts, 2005), but still more than three times the background rate of natural mortality estimated for tropical forests (Brienen et al., 2015; Clark et al., 2004). We used detailed spatial and temporal logging records from Rondônia, Brazil, together with Landsat data, to build a machine learning algorithm for detecting selectively logged Landsat pixels. Machine learning (neural networks, decision trees, support vector machines, etc.) for classification of satellite imagery has been used with increasing success in recent years (Yuuta et al., 2011) and can turn a suite of predictor variables weakly correlated with a response into a relatively strong classifier (Breiman, 2001). The successful application of this algorithm to a test site in northern Pará, Brazil, approximately 1500 km from the location of algorithm development, demonstrates that this approach is transferable and can greatly improve existing methods of detecting subtle selective logging activities in the tropics.

2. Study sites and satellite imagery

Data from two test sites in the Brazilian Amazon were used in this study (Fig. 1a). The Jamari site consists of terra firme tropical forest inside the Jamari National Forest, Rondônia, Brazil. The logging concession was subdivided into forest management units (FMUs) that were each approximately 2000 ha (Fig. 1b). Selective logging occurred within a single FMU in each year, at an intensity of approximately 10 m$^3$ ha$^{-1}$ (1–2 trees ha$^{-1}$), beginning at the end of the wet season (roughly June) and continuing through the dry season (until November) from 2011 through 2015. Forest inventory measurements were recorded by trained foresters and included the spatial location of each marketable tree. At the Jamari site, heavy cloud cover typically occurs between October and May, but cloud-free images from Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper (ETM+) and Landsat 8 Operational Land Imager (OLI) were acquired approximately annually for 2008 to 2016 in the intervening dry season (Table 1). Note that the 2012 ETM+ images suffered from missing data as a result of the scanline corrector error and appear striped (Storey et al., 2005). For the analyses, we distinguished “early” and “late” images for a given region. The early image was the last cloud-free image of the dry season in the same year the FMU was logged (typically in August, approximately 2–3 months before cessation of logging activities for the season). The late image was the first cloud-free image of the dry season in the year after cessation of logging activities (typically in June, approximately 8–12 months after the FMU was logged). We used early and late imagery to generate two separate datasets and build two separate algorithms in order to assess which time period provided better detection of selective logging. This is illustrated for a hypothetical logging season in Fig. 2. The selection of two time periods reflects the fact that after 8–12 months, regrowth of foliage and other vegetation can reduce the spectral signatures required to identify canopy gaps and woody debris in tropical systems (Asner et al., 2004a, b; Broadbent et al., 2006).
The Jari site (Fig. 1c) in Pará, Brazil, consists of terra firme tropical forest inside the 12,500 ha Jari concession that was selectively logged at an intensity of approximately 12 m$^3$ ha$^{-1}$ (1–3 trees ha$^{-1}$) between July and December 2012. In contrast to Jamari, the Jari site lacked detailed information on where trees were removed, but the volume of wood (m$^3$) harvested was recorded for 10 ha (400 m × 250 m) blocks in the concession. The Jari site allowed us to assess whether the algorithms developed using the Jamari dataset, located approximately 1500 km away, were transferable to this distant site. At Jari heavy cloud cover is common throughout the year, but we used the early and late time period imagery with the lowest cloud cover available to assess logging before and after logging activities occurred within the FMU (Table 1).

3. Methods

3.1. Data inputs for detecting selective logging

For the Landsat scenes given in Table 1, the surface reflectance values for the Blue, Green, Red, Near Infrared, Shortwave Infrared 1 and Shortwave Infrared 2 bands were measured at each pixel where logging occurred ($n = 13,699$) and 2000 randomly selected pixels in an adjacent FMU that remained unlogged. In addition, since logging activities tend to be accompanied by surrounding disturbances (residual damage to neighbouring unharvested trees and skid trails along which logs are extracted), seven texture measures were calculated for each band (mean, variance, homogeneity, contrast, dissimilarity, entropy, and second moment) to provide a local context for each pixel (Beekhuizen and Clarke, 2010; Castillo-Santiago et al., 2010; Haralick et al., 1973; Rodriguez-Galiano et al., 2012). These were calculated within a $7 \times 7$ pixel window, chosen as a trade-off between minimizing window size while still capturing the disturbances in a selectively logged forest compared to an unlogged forest (see Section 4.1 for a brief comparison of larger and smaller window sizes). The various texture metrics were assigned to the centre pixel, thus maintaining pixel size (i.e. 30 m), and were added after preliminary modelling efforts with only the surface reflectance bands were found to perform inadequately (i.e. approximately double the rate of omission error of logged pixels; see Table S1 for details). Because of possible Landsat inter-sensor differences, we added one final categorical variable that represented the
that detection of selective logging was better with early time period data.

### 3.2. Random Forest for detection of selective logging

We built Random Forest (RF) models using the `randomForest` package in program R version 3.3.1 (Liaw and Wiener, 2002; R Development Core Team, 2016). The RF algorithm (Breiman, 2001) is a machine learning technique that uses an ensemble method to identify a response variable (here, whether a pixel was logged or unlogged) given a set of predictor variables (e.g. surface reflectance values). In contrast to a single decision tree, RF models employ multiple, independent decision trees (hence a forest). Random subsets of the training data are drawn, with replacement, to construct many trees in parallel, with each tree casting a vote on which class should be assigned to the input data. The withheld subset of the data, called the out-of-bag fraction, can be used for validation in the absence of independent validation data (Breiman, 2001). To reduce generalization error, RF also uses a random subset of predictor variables in the decision at each node within a tree during construction.

We split the early and late datasets into 75% for training and 25% was withheld for validation. We used the out-of-bag data during model training to determine the threshold value for classification (i.e. model calibration, see Section 3.3.1). In order to ensure independence, the training and validation datasets were spatially filtered such that no observations in the training dataset were within 90 m of an observation in the validation dataset. RF models have only two tuning parameters: the number of classification trees to be produced (\(k\)), and the number of predictor variables used at each node (\(m\)). We used 10-fold cross-validation to identify the number of trees (\(k = 1000\)) and the number of variables to use at each node (\(m = 5\)) that minimized the out-of-bag error rate on the training data.

#### 3.3. Algorithm evaluation

**3.3.1. Calibration: selecting the detection threshold**

RF models typically use a simple majority vote to assign an observation to a particular class, for example, in binary decisions when > 50% of the trees assign a pixel to a particular class (Breiman, 2001). However, the proportion of votes cast for a particular class from the total set of trees can be obtained for each pixel and a classification threshold can be applied to this proportion (Liaw and Wiener, 2002). We adopted this approach here, wherein the proportion of votes that predicted each observation to be logged, denoted as \(X\) and informally termed the likelihood a pixel was logged, was used to select the classification threshold. Model calibration (with the out-of-bag data) was then used to define a threshold, \(T\), such that if \(X > T\) the pixel was classified as logged (Fig. 3).

Detection of logging involves only two classes, logged and unlogged forest, so the confusion matrix has the form:

---

**Table 1**

| Study site | Acquisition date | Scene timing | Solar zenith angle | Landsat sensor |
|------------|------------------|--------------|--------------------|----------------|
| Jamari     | 2008-07-28       | Early        | 49.75              | TM             |
|            | 2009-07-31       | Early        | 50.00              | TM             |
|            | 2010-07-18       | Early        | 46.36              | TM             |
|            | 2011-08-06       | Early        | 51.67              | TM             |
|            | 2012-08-16       | Early        | 54.05              | ETM +          |
|            | 2013-08-27       | Early        | 57.07              | ETM +          |
|            | 2014-08-30       | Early        | 58.84              | OLI            |
|            | 2015-09-02       | Early        | 60.19              | OLI            |
|            | 2009-06-29       | Late         | 43.79              | TM             |
|            | 2010-07-02       | Late         | 43.63              | TM             |
|            | 2011-07-05       | Late         | 44.30              | TM             |
|            | 2012-06-13       | Late         | 41.64              | ETM +          |
|            | 2013-07-10       | Late         | 42.26              | OLI            |
|            | 2014-06-11       | Late         | 40.43              | OLI            |
|            | 2015-06-14       | Late         | 40.47              | OLI            |
|            | 2016-06-16       | Late         | 40.37              | OLI            |
| Jari       | 2011-11-08       | Early        | 123.31             | ETM +          |
|            | 2012-11-10       | Early        | 125.27             | ETM +          |
|            | 2011-07-03       | Late         | 48.12              | ETM +          |
|            | 2013-08-17       | Late         | 60.92              | OLI            |

---

**Fig. 2.** Timeline representation of a single forest management unit in the Jamari study site. Vertical blue lines indicate image acquisitions during the early and late time periods (black boxes) relative to when logging occurred (red box). In this example the early Landsat image was acquired part way through the logging season, so part of the management unit has yet to be cut. The late image is the first cloud-free image of the following dry season and is acquired approximately 8 months after the management unit was selectively logged. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
where \( L \) and \( UL \) refer to logged and unlogged, \( N_L \) and \( N_{UL} \) are the numbers of logged and unlogged observations in the reference dataset, and \( D_L \) and \( D_{UL} \) are respectively the numbers of logged and unlogged pixels detected as logged. The total number of observations is \( N = N_L + N_{UL} \). Since logging is a relatively rare event, both in our data and on the landscape (i.e., \( N_L \ll N_{UL} \)), it is appropriate to use the terminology of detection theory. Accordingly, we define the detection probability \( P_d = D_L/N_L \) and false detection probability \( P_{fd} = D_{UL}/N_{UL} \) as the probabilities that a logged or unlogged pixel is classified as logged, respectively. \( P_d \) is equivalent to \( 1 - \) the omission error of the logged class and \( P_{fd} \) is the omission error of the unlogged class.

A pixel was classified as logged if \( X \), the proportion of votes from RF that predict the pixel as logged, exceeds a given threshold \( T \). Hence the detection and false detection probabilities depend on \( T \) and can be written

\[
P_d(T) = \int_T^1 f_L(X) \, dX 
\]

and

\[
P_{fd}(T) = \int_0^T f_{UL}(X) \, dX 
\]

where \( f_L(X) \) and \( f_{UL}(X) \) are the probability distributions of \( X \) for the logged and unlogged classes, respectively (see Fig. 3).

The selection of \( T \) involves a trade-off between increasing \( P_d \) and reducing \( P_{fd} \) (Fig. 3). In making this choice, the overall accuracy, given by

\[
A = \frac{D_L + (N_L - D_L)}{N},
\]

is not a good guide, since it can be shown that \( A \) is maximal (equivalently, the overall probability of error is a minimum) when

\[
\frac{f_L(X)}{f_{UL}(X)} = \frac{N_{UL}}{N_L}. 
\]

If \( N_L \) and \( N_{UL} \) were equal, the threshold would then be chosen at the intersection of \( f_L(X) \) and \( f_{UL}(X) \), but since \( N_L \ll N_{UL} \) it has a much higher value (i.e., it moves to the right in Fig. 3). This is because to increase overall accuracy it is more effective to reduce \( P_{fd} \) than to increase \( P_d \), since there are so many more unlogged pixels (Schwartz, 1984), and maximizing accuracy would lead to very few (or even no) detections. For example, if only 1% of an area was logged and all the pixels were classified as unlogged, the overall accuracy would be 99%. Thus, overall accuracy would not sufficiently balance the trade-off between true and false detections to meet our objectives.

Various criteria could be used to select a classification threshold, including maximizing Cohen’s kappa (Cohen, 1960) or defining an acceptable rate of omission error; ultimately however, there is no wrong threshold, since this depends on the objectives of prediction. The criterion used in this study to define \( T \) was to fix the proportion of detected pixels that were truly logged, defined here as \( d_{pl} \):

\[
d_{pl} = \frac{D_L}{D_L + D_{UL}} = 1 + \frac{1}{\left( \frac{N_{UL}}{N_L} \right) \left( \frac{P_d}{P_{fd}} \right)} 
\]

Adopting this criterion is equivalent to a Constant False Discovery Rate detector which is widely used in detection problems with rare events (Benjamini and Hochberg, 1995; Neuvial and Roquain, 2011). This fixes the rate of prediction error (i.e., type I) when labelling pixels as logged, because \( d_{pl} \) is equal to 1 minus the commission error of the logged class, thus limiting the rate of commission error. This approach enables the user to select the proportion of detections that will be false. It was chosen because in the detection of rare events (e.g. selective logging within the Amazon Basin, for example), the implications of a particular error rate when predicting over the majority class (i.e., unlogged forest) are greater than an equivalent error rate when predicting over the minority class (i.e., 10% of millions of unlogged pixels > 10% of thousands of selectively logged pixels). Thus, in order to avoid being swamped by false detections, we wanted to fix the proportion of all detected pixels that were incorrect and accept the level of accuracy associated with this criterion. The approach outlined here, therefore, should be viewed from a detection theory perspective as opposed to simply being a classification problem.

Model calibration was used to calculate \( P_{pl}, P_{fd}, \) and \( d_{pl} \) across the full range of threshold values. In practice this involved iterating through all values of \( T \) between 0 and 1 (in steps of 0.001), building each confusion matrix, and calculating the associated values of \( P_d, P_{fd}, \) and \( d_{pl} \). The threshold value was chosen such that \( d_{pl} = 0.85 \) in the training data (i.e. 15% of pixels classified as logged were actually unlogged). We initially set \( d_{pl} \) to 95% to strongly limit the rate of false
were used to assess agreement and calculate detection rates, but this resulted in very high omission error of truly logged pixels (> 75%). Consequently, \(d_{\text{om}}\) was reduced to 0.85 by lowering the threshold, thus causing the detection and false detection rates to increase and causing more logged pixels to be detected. This value was then used to estimate \(P_d\) and \(P_{fd}\) during model assessment with the validation dataset.

### 3.3.2. Validation: assessing model accuracy

RF models were validated using a random, independent subset of the early and late datasets (described in Section 3.2). The threshold value of \(T\), chosen during model calibration, was applied to the validation data and the associated error rates were calculated. The values of \(P_d\), \(P_{fd}\) and \(d_{\text{om}}\) are presented across full range of threshold values to thoroughly illustrate model performance. Good practices outlined by Olofsson et al. (2014) were used to assess agreement and calculate unbiased error estimates when mapping selective logging detections. During mapping, non-forested areas were excluded using Brazil’s national forest change product, PRODES (INPE, 2015), and cloudy pixels were masked using the cloud mask provided with Landsat surface reflectance imagery. In addition, we provide the value of Cohen’s kappa, \(\kappa\), for comparison with other studies (Cohen, 1960).

### 4. Results

#### 4.1. Random Forest classification of selective logging at Jamari

The rates of true and false detection probabilities for the early and late validation data are shown in Fig. 4 for the full range of \(T\) (black lines). These curves indicate how a given threshold value used for classification influenced the associated values of \(P_d\), \(P_{fd}\), \(\kappa\), and \(d_{\text{om}}\) in the validation data. For example, if a \(d_{\text{om}}\) of 0.90 was used (indicating 10% of logging detections would be spurious) then the false detection rate \(P_{fd}\) would be < 1% for both datasets, but the detection rate \(P_d\) would be approximately 55% and 30% for the early and late datasets, respectively. These plots clearly demonstrate that there is no ambiguous way to choose an optimal value for \(T\), and the choice about its value is a trade-off between the number of true and false detections.

In general, these plots indicate that the early data provided a higher detection rate than the late data, for a given false detection rate. The early and late data had similar rates of commission error when labelling logged pixels, which is not surprising given we used this measure to constrain models during training. However, the late data had higher rates of omission error of logged pixels and detected less logging (Table 2). In addition, these plots demonstrate why using the threshold that maximized Cohen’s \(\kappa\) would lead to higher false detection rates, as the threshold value is higher when \(d_{\text{om}}\) is 0.85 than at maximum \(\kappa\) (i.e. pixels classified as logged must have a higher likelihood). Furthermore, because \(\kappa\) is high across a wide range of threshold values for both early and late data, slight differences in the likelihoods produced by the validation data could result in dramatic shifts in the value of \(T\).

Although \(d_{\text{om}}\) was fixed at 0.85 during model calibration (i.e. with the training data), the values calculated with the validation dataset were slightly lower (Table 2). Thus, the threshold value determined during model training did not produce the same values for \(d_{\text{om}}\) when used against the validation dataset (i.e. some loss of performance). Slight differences in the proportion of logged observations (16.3% and 14.5% in training and validation, respectively) and minor differences in the ratio of \(P_{fd}\):\(P_d\) between the training and validation datasets account for the disparity (see Eq. (4)). In general model assessments seldom give identical performance across training and validation phases, and the difference here were marginal and yielded comparable model behavior.

The early data displayed higher spatial correspondence between high likelihoods and the locations of logging in Jamari. This is illustrated in Fig. 5, where the likelihood of logging provided by RF is shown on a color scale and the individual locations of tree removal are indicated by black squares. The early model yields much higher likelihoods and these match well with reference logging data, whereas there is generally lower correspondence between reference logging locations and regions of highest likelihood in the late predictions. Note that we expect some logging locations to be omitted in the early data as the corresponding satellite data were acquired part way through the logging period and missed later logging. Evidence for this is provided by the inset regions expanded at the bottom of Fig. 5 where the locations of the last 200 trees in the logging records for the season are displayed as crosses instead of squares. Many of these locations occur in low likelihood regions in the early data because these locations were probably unlogged at the time of the image acquisition (dates for specific tree removal were unavailable).

A further marked difference between the predictions is that, in general, far more pixels were labelled as logged in the early data than in the late, as can be seen by comparing the classifications in Fig. 6, which shows the years between 2011 and 2015 for the early (top) and late (bottom) datasets, respectively. The FMU where logging occurred in each year is outlined in yellow and the 2015 image also shows the FMU to be logged in 2016 outlined in white. The early classifications appear to show some indication of a retained signal from the previously logged FMU that is not as apparent in the later data, which resulted in some scenes having very few pixels of high likelihood of logging (see 2012-06-13 in Fig. 6) and others with most of the study area predicted as logged (see 2016-06-16 in Fig. 6). This suggests the threshold value from model calibration could not be used reliably for all late images and a scene-specific threshold value might need to be calculated for each image to provide better correspondence with logging activities.
The true proportion of logged pixels in each FMU (from the logging records) was roughly 12% in a given year (mean = 11.8%; standard deviation = 2.4%), but the early classifications consistently labelled a greater number of pixels as logged (Fig. 7). For example, the proportion of pixels assigned in each FMU for early acquisitions was expected to be around 25% (10% truly logged and 15% false positives), but nearly

### Table 2

Confusion matrix summarizing unbiased (Olofsson et al., 2014) results from Random Forest (RF) model classifications of logged and unlogged observations at Jamari derived from Landsat data at labelled points (observations before and after selective logging). The classification threshold ($T$) for RF models was set during model calibration such that the proportion of detections that were truly logged ($d_{pl}$) was fixed at 0.85, resulting in a $T$ of 0.40 and 0.65 for the early and late datasets, respectively. The corresponding values for overall accuracy (OA), Cohen's kappa ($\kappa$), the proportion of detected pixels that were truly logged ($d_{pl}$), and the detection probability ($P_d$) are provided.

|            | Early |          | Late |          |
|------------|-------|----------|------|----------|
|            | OA: 89.7% | $\kappa$: 0.78 | OA: 91.7% | $\kappa$: 0.40 |
|            | $d_{pl}$: 0.80 | $P_d$: 0.92 | $d_{pl}$: 0.80 | $P_d$: 0.30 |
| Predicted Class | Logged | Unlogged | Error (%) | Logged | Unlogged | Error (%) |
|               | 0.313 | 0.076 | 19.5 | 0.032 | 0.008 | 19.9 |
| Omission Error (%) | 8.0 | 11.5 | | 70.1 | 1.0 | |

The true proportion of logged pixels in each FMU (from the logging records) was roughly 12% in a given year (mean = 11.8%; standard deviation = 2.4%), but the early classifications consistently labelled a greater number of pixels as logged (Fig. 7). For example, the proportion of pixels assigned in each FMU for early acquisitions was expected to be around 25% (10% truly logged and 15% false positives), but nearly

---

**Fig. 5.** Example of a forest management unit in Jamari logged in 2013 showing the RF predicted likelihood that each pixel was logged (highest likelihoods in red) for the early and late data. Logging roads are thin black lines and tree removal locations are displayed as black squares and crosses. The black crosses (see insets for detail) coincide with the final 200 trees in the logging records for 2013. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
twice as many were identified. However, forest disturbances from selective logging affect patches of forest and not just the pixels where trees were logged. Extra detections would be expected because of additional tree and canopy damage associated with tree removals, roads, and construction of skid trails. Note that the rate of false detections over unlogged FMUs (open diamonds in Fig. 7) is roughly as expected for the early algorithm and most dates for the late algorithm, but is significantly different for the late algorithm for the FMU logged in 2015. The late scene for this FMU clearly shows anomalous behaviour and displays high likelihood of logging over most of the study area, including known unlogged regions (see Fig. 6).

We used the early algorithm to predict over the available Landsat time series in Jamari that coincided with logging in four FMUs (see Table S3 for image dates) and plotted the detections of logging through time (Fig. 8). As expected, the proportion of detected pixels increased through the logging season during the year a given FMU was logged. There was also a drift upwards in the unlogged FMU, but the detections peaked just above the expected rate of 12% by late August (Fig. 8). Importantly, known unlogged regions will not exhibit a $d_{PL}$ of 0.85 (i.e. a false discovery rate of 15%), as any and all detections in known unlogged areas are wrong (i.e. $d_{PL} = 0$). Consequently, the false alarm rate is the expected proportion of detections (i.e. $P_{FA} = 11.5\%$ in Table 2). This suggests that the algorithm performed as would be expected for tracking forest disturbances through time in both logged and unlogged FMUs. In particular, forest patches subjected to selective logging should display measurable increases in detections as the logging season progresses and known unlogged regions will exhibit the expected false alarm rate.

We assessed the impact of the window size used to calculate texture measures on the proportion of pixels labelled as logged in four logged and one unlogged FMU in the early data (Fig. 9). Reducing the

---

**Fig. 6.** Classifications for Jamari between 2011 and 2016 with early (top) and late (bottom) Landsat data. The forest management units (FMUs) are outlined in black and the FMU logged in each year (where logging should be detected) is outlined in yellow. Blue and green represent classifications for logged and unlogged forest, respectively. White areas are no-date and correspond to the Landsat 7 scan-line corrector error (stripes) and pixels that were non-forest (irregular patches) in Brazil’s Program to Calculate Deforestation in the Amazon (PRODES) database. The FMU logged in 2016 is outlined in white (far right) and the top two FMUs in each image remained unlogged. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. 7.** The proportion of pixels in each FMU that were classified as logged in Fig. 6 for the early (open symbols) and late (closed symbols) algorithms. Circles are the logged FMUs in each year and diamonds are values from an FMU that remained unlogged. The black line represents the mean ± 1 standard deviation (dashed lines) of the true rate of logging across all FMUs. Values are unbiased (Olofsson et al., 2014) to account for possible sampling bias in the validation data.

**Fig. 8.** The proportion of pixels classified as logged through time in three logged and one unlogged FMU using the early RF model. Triangles, circles, and squares represent logged FMUs (solid lines) and diamonds are an unlogged FMU (dotted line). The grey horizontal line at 12% is the approximate detection rate expected for unlogged regions. Values are unbiased (Olofsson et al., 2014) to account for possible sampling bias in the validation data.
window size from $7 \times 7$ to $3 \times 3$ lowered the proportion labelled as logged by nearly 50% within each FMU, resulting in smaller clusters of pixels with high likelihoods (Fig. 9). However, as noted above, forest disturbance from selective logging affects chunks of forest and not just the pixels where trees are cut. Thus, depending on the scale of interest, larger or smaller window sizes may be better for identifying patches of forest that have been selectively logged. In contrast, reducing the window size had little impact on the false detection rate over unlogged regions, remaining close to the 12% expected irrespective of window size (Fig. 9). This suggests that the choice of window size is independent of the false positive rate over undisturbed forested areas and primarily affects likelihoods around pixels that the algorithm identifies as disturbed.

### 4.2. Random Forest predictions of logging at Jamari

The majority of the best available (sufficiently cloud-free) Landsat scenes over Jamari were from the ETM+ sensor, which suffered the scanline corrector error, so approximately 22% of each image has missing data that appear as white stripes in Figs. 10 and 11 (Storey et al., 2005). Nonetheless, this allowed us to see behaviour similar to Jamari, wherein predictions using early data clearly identify active logging (Fig. 10) and predictions using late data detected very little logging (Fig. 11). In particular, with late data most of the study area was classified as unlogged both before and after logging. Additionally, with early data the predictions of logged pixels in the year before logging were close to the expected rate of false positives over unlogged regions (approximately 12%). However, with late data the rate of false positives was not close to the expected rate over unlogged regions. Maps for the year before logging are displayed to demonstrate that the early dataset identified the correct year in which logging occurred and did not simply predict high amounts of logging for every year.

In total, an area of 6152 ha was visible in Jamari after removing clouds and missing data gaps from the SLC error in the year of logging. Of this area, 1710 ha was not logged (black boxes in Figs. 10 and 11). Since we lacked detailed logging records and only knew which 10 ha blocks were logged, a formal accuracy assessment of logging detections was not possible. However, when using the unbiased proportions and the threshold from Table 3 to classify predictions, the early algorithm labelled 2316 ha (38%) as logged (Fig. 10). This value is consistent with predictions from Jamari where approximately 40% of logged FMUs were labelled with early data (see Fig. 7). In addition, the rate of commission error when predicting logged pixels (i.e. $1 - d_{cm}$) was 19.8%, which is also consistent with the rate of commission error between the validation data and prediction errors found for the Jamari site.

### 5. Discussion

The spatial resolution of Landsat data has previously been considered too coarse to monitor selective logging activities (Asner et al., 2002), with most applications involving logging intensities $> 20$ m$^3$ ha$^{-1}$ at sites with an abundance of spectrally distinct features (Asner et al., 2005; Souza and Barreto, 2000; Souza et al., 2005). However, we have demonstrated that Landsat surface reflectance data can be used effectively, in a supervised machine learning framework, to detect subtle spectral changes from selective logging at low intensities. Although a definitive estimate of the amount of logging activities that have previously gone undetected is difficult to determine, a dataset of 824 logging permits from the state of Pará, Brazil found 18% of permits authorized for logging were harvested at intensities $< 20$ m$^3$ ha$^{-1}$ (Richardson and Peres, 2016). Thus, our approach has the potential to significantly increase current abilities to detect and monitor selective logging activities that up to now have been, at best, marginally detectable (see Supplementary material, Section 3 for a comparison between our method and CLASlite, Asner et al., 2009a). In addition, the approach outlined here has the distinct advantage of being able to make predictions about forests on a single scene to map disturbances, instead of requiring successive cloud-free images like many approaches (Asner et al., 2009a). This is particularly important since a single, low-cloud scene may be all that is available for a given region (see Souza Jr et al., 2013).

Only the algorithm developed with data close to the time of active logging (i.e. the early data) performed well at detecting selective logging. Many logged pixels were omitted when using data from the first cloud-free image of the next dry season (i.e. late). In addition, only the algorithm trained with imagery close in time to the logging events was transferable to new areas (Figs. 10 and 11). Thus, our results suggest images acquired during, or very soon after, active logging are needed to map low intensity selective logging. This is partly because logging activities typically occur in the dry season when cloud-free imagery is more likely to be available, but also because the spectral changes associated with low-intensity selective logging practices are subtle and short-lived and rapidly become obscured under even limited regrowth (Broadbent et al., 2006).

The decision to fix the proportion of logging detections that were correct (i.e. limiting the commission error when predicting logged pixels) defined the classification threshold applied to the likelihoods produced by the RF models developed at Jamari. This threshold would likely give different values of $d_{cm}$ in regions that contain different proportions of logged and unlogged observations (see Eq. (4)). Indeed, the threshold value from model training produced a slightly higher $d_{cm}$ when assessed against the validation dataset, yet these data were from the same study site. In addition, depending on the distribution of likelihoods produced by the RF models, different datasets might yield different threshold values, for example because of higher selective logging intensities. However, assuming both classes are present, the proportion of detected pixels that are wrong (i.e. $1 - d_{cm}$) would be expected to remain invariant. Hence if the same threshold were applied over the whole of the Amazon basin, we would expect approximately 20% of all detections to be wrong and 11.5% of truly intact forest pixels to be identified as logged. This could be used to refine the algorithm (in the absence of field data on logging locations) by examining the rate of false detections over known unlogged regions or protected areas to achieve a similar error rate. Adopting this threshold (i.e. $d_{cm} = 11.5$) would make the method equivalent to a Constant False Alarm Rate detector which is widely used in detection problems with rare events (Scharf, 1991). A $d_{cm}$ of 85% was the value chosen here as a compromise that gives a high detection rate (0.92 for early data, see Table 2) while keeping the proportion of detections that are false to an acceptable
level. However, other values of \( d_{pl} \) could be chosen, depending on the predictive objectives of the particular application. This is precisely why Fig. 4 shows the full range of threshold values; to enable a detailed assessment of model performance with higher or lower values of \( T \) or \( d_{pl} \).

An important issue when assessing detections of selective logging is that patches of forest are affected, not just the isolated pixels where trees are removed. The area around logged pixels is certain to be disturbed because of canopy damage associated with tree removals and the construction of roads and skid trails, but the precise amount is unknown. Consequently, taking as a reference purely the pixels where trees were known to be removed is inadequate for assessing the disturbance due to logging. Indeed, the true rate of logged pixels at Jamari was approximately 12% (mean = 11.8%; standard deviation = 2.4%), but this represents a minimum expected detection rate and the associated forest disturbances would result in more detections. The early algorithm labelled approximately 40% of the area inside FMUs in Jamari and Jari as logged. This may be a more realistic estimate and is likely close to the upper limit of what constitutes forest disturbance for this level of logging. However, because the choice of window size for texture measure calculation affected the proportion of pixels labelled as logged (Fig. 9), the appropriate window size for a particular application needs to be considered. Smaller windows resulted in fewer detections, but use of too small a window risks being unable to adequately measure texture arising from forest disturbances from selective logging. Thus, the specific application would best dictate the optimum approach and the user should, if possible, use window sizes matched to the expected or known spatial spread of forest disturbance around tree removals.

Selective logging rates in the Brazilian Legal Amazon (BLA) are thought to have remained relatively stable since 2000, with Pará and Mato Grosso enduring the highest rates of selective logging (Betts et al., 2017; Souza Jr et al., 2013). However, our findings suggest that their assessments of forest disturbance and the associated carbon emissions are likely underestimated. Machine learning approaches (neural networks, decision trees, support vector machines, etc.) for classification of satellite imagery have been used with increasing frequency and success since their initial applications to remote sensing questions in the 1990’s (Tuia et al., 2011), but their effectiveness relies heavily on adequate training data. Our results suggest that detailed logging records ought to be a reporting requirement for logging companies or for REDD+ projects related to logging. These datasets could be used for building, improving, and updating models similar to the one presented here, with the aim of facilitating the creation of pan-tropical estimates of (legal and illegal) selective logging activities.

Fig. 10. Logged (blue) and unlogged (green) predictions at the Jari study site using a Random Forest model trained from early Landsat inputs. Predictions from November 2011 (top) were before logging activities began and from November 2012 (bottom) while active logging was ongoing. Clouds were masked out and appear as irregular white patches (top). Missing data regions from the Landsat 7 scan-line corrector error appear as white stripes through the maps. Black boxes indicate the 10 ha blocks inside the Jari concession that were not logged. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
From a conservation perspective, the ability to identify regions of forest that are selectively logged is useful for mapping primary forest, but also for delineating logged forests with conservation value. Forests subjected to selective logging generally maintain far higher levels of biodiversity than other modified habitats, such as plantations or secondary forests (Gibson et al., 2011; Edwards et al., 2014). Moreover, even after accounting for the amount of wood removed, reduced-impact logging activities (like those at our study site in Jamari) do better at maintaining biodiversity than conventional selective logging practices (Bicknell et al., 2014) while simultaneously sequestering more carbon during regrowth (Martin et al., 2015; Putz et al., 2008). Thus, in the context of REDD+ or alternative conservation initiatives, forests affected by low intensity selective logging offer high biodiversity value and carbon sequestration potential. Accordingly, our method could be

Table 3
Confusion matrix summarizing unbiased (Olofsson et al., 2014) results from Random Forest (RF) model classifications of logged and unlogged observations at Jari with Landsat data. The thresholds (T) developed at Jamari were used to classify predictions at Jari and were 0.40 and 0.65 for the early and late datasets, respectively (Table 2). The corresponding values for overall accuracy (OA), Cohen’s kappa (κ), the proportion of detected pixels that were truly logged ($d_{pL}$), and the detection probability ($P_d$) are provided.

|                | Reference Class | Commission |
|----------------|-----------------|------------|
|                | Logged (OA)     | Unlogged   | Error (%) |
| Early          | 0.351           | 0.085      | 19.5      |
|                | 0.025           | 0.538      | 4.4       |
| Omission Error | 6.7             | 13.7       |           |
| Late           | 0.002           | 0.005      | 19.9      |
|                | 0.078           | 0.919      | 7.8       |
| Omission Error | 97.3            | 0.06       |           |
used for identifying and prioritizing forest tracts suitable for such initiatives.

5.1. Study limitations

While the minimum mapping unit remained 30 m, the use of texture measures resulted in some spatial aggregation of logging predictions (see Figs. 5 and 6). This was expected around logged pixels, as a result of canopy gaps, skid trails, and roads, but clustered detections were also present in unlogged FMUs (see Fig. 6). Ideally, predictions of logging in unlogged FMUs would have shown a diffuse 12–15% of spurious detections. Attempts to refine the accuracy of a final predictive map, by performing a post-processing step in which either likelihoods or classified pixels are re-examined (e.g. using a window analysis to apply neighbourhood rules whereby likelihoods or counts of nearby pixels are re-evaluated against some criteria) to enhance the detection rate or limit the false detection rate further, would prove difficult (Huang et al., 2014). However, using a smaller window size for texture calculation, such as 5 × 5 pixels, would reduce this effect. Ultimately, the optimal window size for textures depends on the objectives of the application and understanding how different window sizes affect detection and false detection rates.

Landsat surface reflectance data is known to exhibit occasional strong scene-to-scene and within-scene variations because of discontinuities across focal plane modules (Morfit et al., 2015) and seasonal changes in solar viewing angles (Roy et al., 2016), respectively. We did not take these effects into account and likely affected algorithm performance in some instances (e.g. 2016-06-16 in Fig. 6). Thus, a large scale application of the approach outlined here should include a step to normalize surface reflectance data across scenes to facilitate detection of the subtle and short lived spectral changes associated with low-intensity selective logging practices (Broadbent et al., 2006).

Our analysis used a binary classification (logged and unlogged forest) yet tropical forest landscapes are a heterogeneous mixture of land uses (e.g. secondary forests, burned areas inside forests, agricultural fields). We avoided some of these complexities by using the PRODES forest designations to remove urban areas, agricultural fields, and deforested areas that had regenerated to secondary forest. However, our method cannot distinguish between disturbance types and is best suited for tracts of remaining forest that contain logging concessions. In addition, selective logging represents a range of forest disturbance intensities and we would have preferred to use the logging dataset in a regression framework (i.e. a continuous response, such as logging intensity). However, the range of logging intensities within our Jamari dataset was very limited, since it was such a low intensity concession. Consequently, a regression approach was not suitable for the Jamari dataset and we chose to use classification. Additional datasets could fold into the framework here and might facilitate a continuous response approach as those datasets become available.

Finally, our analyses used freely available optical datasets. However, the problems associated with using optical imagery in the tropics, including the limited availability of cloud-free images over many regions and the rapid regeneration of tropical forest vegetation, remain major obstacles to pan-tropical assessments of selective logging rates. Methods that integrate optical and radar dataset into a single algorithm would likely further improve the detection of tropical selective logging activities (Higginbottom et al., 2018; Joshi et al., 2016; Reiche et al., 2018).

6. Conclusion

Loss and degradation of forests in the tropics has important implications for global climate change, local populations and biodiversity (Lewis et al., 2015). Methods to reliably map forest disturbances from selective logging would be a key contribution to quantifying the terrestrial portion of the carbon budget and the role of land-use change in tropical forests emissions (Baccini et al., 2017). In addition, reliable forest monitoring systems are actively sought after by tropical nations and conservation groups tasked with mitigating global climate change through improved forest management practices (GOFC-GOLD, 2016). Our results should stimulate further assessments of regional rates of low-intensity selective logging in tropical forests.

Our analysis, based on training Random Forest models with detailed records of tree removals, has demonstrated that Landsat data can be effective at detecting selective logging at much lower intensities than has previously been reported. To be successful, the input satellite data needs to be acquired within a few months of the logging, as the subtle signal caused by logging (and the more extensive disturbance associated with logging) is rapidly lost. Although we had less complete knowledge of logging activities at the Jari site, the algorithm developed at Jamari appeared to transfer successfully to this site (despite being 1500 km away). Hence there is reason to expect that it could be applied at much wider scales.

Acknowledgements

We would like to thank AMATA Brazil for providing access to logging records for Jamari and Jari Florestal for logistical support. MGH was funded by the Grantham Centre for Sustainable Futures. JMBC was funded as part of NERC's support of the National Centre for Earth Observation. FMW was CNPq and NERC-funded (NE/P004512/1; PELD-RAS 441659/2016-0, respectively) and was funded by CAPES (BEX5528/13-5) and CNPq-PELD site 23 (403811/2012-0) during the long-term monitoring in Jari. We thank three anonymous reviewers for their thoughtful comments and suggestions that greatly improved the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2018.11.044.

References

Achard, F., DeFries, R., Eva, H., Hansen, M., Mayaux, P., Sibig, H., 2007. Pan-tropical monitoring of deforestation. Environ. Res. Lett. 2, 45022. https://doi.org/10.1088/1748-9326/2/4/045022.

Alamgir, M., Campbell, M.J., Sloan, S., Gouseem, M., Clements, G.R., Mahmoud, M.I., Houghton, R.A., 2014. Economic, socio-political and environmental risks of road development in the tropics. Curr. Biol. 24, R1130–R1140. https://doi.org/10.1016/j.cub.2014.03.067.

Alamer, G.P., Keller, M., Pereira, R., Zweede, J.C., 2002.Remote sensing of selective logging in Amazonia. Remote Sens. Environ. 80, 483–496. https://doi.org/10.1016/S0034-4257(01)00326-1.

Alamer, G.P., Keller, M., Pereira, J.R., Zweede, J.C., Silva, J.N.M., 2004a. Canopy damage and recovery after selective logging in Amazonia: field and satellite studies. Ecol. Appl. 14, 280–298. https://doi.org/10.1890/01-6019.

Alamer, G.P., Keller, M., Silva, J.N.M., 2004b. Spatial and temporal dynamics of forest canopy gaps following selective logging in the eastern Amazon. Glob. Chang. Biol. 10, 765–783. https://doi.org/10.1111/j.1365-2486.2003.00756.x.

Ammer, G.P., Knapp, D.E., Broadbent, E.N., Oliveira, P.J.C., Keller, M., Silva, J.N., 2005. Selective logging in the Brazilian Amazon. Science 310, 480–482. https://doi.org/10.1126/science.1118051.

Ammer, G.P., Knapp, D.E., Balaji, A., Piáez-Acosta, G., 2009a. Automated mapping of tropical deforestation and forest degradation: CLASlite. J. Appl. Remote. Sens. 3, 33543. https://doi.org/10.1364/JOSAD.3.0032675.

Ammer, G.P., Rude, T.K., Aide, T.M., DeFries, R., Emerson, R., 2009b. A contemporary assessment of change in humid tropical forests. Conserv. Biol. 23, 1386–1395. https://doi.org/10.1111/j.1523-1739.2009.01333.x.

Baccini, A., Walker, W., Carvalho, L., Farina, M., Sulla-Menashe, D., Houghton, R.A., 2017. Tropical forests are a net carbon source based on aboveground measurements of gain and loss. Science 358, 230–234. https://doi.org/10.1126/science.aam9562.

Beekhuizen, J., Clarke, K.C., 2010. Toward accountable land use mapping: using geo-computation to improve classification accuracy and reveal uncertainty. Int. J. Appl. Earth Obs. Geoinf. 12, 127–137. https://doi.org/10.1016/j.jag.2010.01.005.

Benjamin, V., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57 (1), 289–300.

Betts, M.G., Wolf, C., Ripple, W.J., Phelan, B., Millers, K.A., Duarte, A., Butchart, S.H.M., Levi, T., 2017. Global forest loss disproportionately erodes biodiversity in intact
DeFries, R.S., Rudel, T., Uriarte, M., Hansen, M., 2010. Deforestation driven by urban expansion: a review and synthesis. Environ. Res. Lett. 5, 34008.

Bicknell, J.E., Struebig, M.J., Edwards, D.P., Davies, Z.G., 2014. Improved timber harvest selectivity in tropical forests through spatially explicit decision support systems. Remote Sens. Environ. 148, 42–57.

Gibbs-Hango, T., Paton, D., Struebig, M.J., Edwards, D.P., Boege, B., 2017. Mapping deforestation using Landsat data. ISPRS J. Photogramm. Remote Sens. 139, 88–102. doi:10.1016/j.isprsjprs.2018.02.010.

Agostini, P., DeFries, R.S., Peh, K.Y., Van Der Sluijs, J., Van Der Laan-Luijkx, T., Ziegler, M., 2016. Forest restoration at large scales: a synthesis of recent trends in tropical forests. Front. For. Global Change 9, 16.

Sanderson, E.B., Scharlemann, J.P.W., Erondu, N.N., 2013. Reducing deforestation and forest degradation from agriculture. BioScience 63, 91–99. doi:10.1093/biosci/bis100.

Kuemmerle, T., Meyfroidt, P., Mitchard, E., Reiche, J., Ryan, C., Waske, B., 2016. A global forest metrics database for monitoring forest disturbance. Remote Sens. Environ. 139, 88–102. doi:10.1016/j.isprsjprs.2018.02.010.

Herold, M., Johns, T. 2007. Linking requirements with capabilities for deforestation monitoring in the context of the UNFFC-REDD process. Environ. Res. Lett. 2, 45025. doi:10.1088/1748-9246/2/4/045025.

Ruckes, C.M.,基础设施, J., 联合, J., 哈里克, K.W., Higginbottom, B.B., 2013. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. doi:10.1016/j.rse.2014.05.006.

Bicknell, J.E., Struebig, M.J., Edwards, D.P., Davies, Z.G., 2014. Improved timber harvest selectivity in tropical forests through spatially explicit decision support systems. Remote Sens. Environ. 148, 42–57.

Herold, M., Johns, T. 2007. Linking requirements with capabilities for deforestation monitoring in the context of the UNFFC-REDD process. Environ. Res. Lett. 2, 45025. doi:10.1088/1748-9246/2/4/045025.

Ruckes, C.M., 基础设施, J., 联合, J., 哈里克, K.W., Higginbottom, B.B., 2013. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. doi:10.1016/j.rse.2014.05.006.

Herold, M., Johns, T. 2007. Linking requirements with capabilities for deforestation monitoring in the context of the UNFFC-REDD process. Environ. Res. Lett. 2, 45025. doi:10.1088/1748-9246/2/4/045025.
V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172. https://doi.org/10.1016/j.rse.2014.02.001.

Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J.I., Lewis, P.E., Schaaf, C.B., Sun, Q., Li, J., Huang, H., Kovalskyy, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sens. Environ. 176, 255–271. https://doi.org/10.1016/j.rse.2016.01.023.

Sanaki, N., Potz, P.E., 2009. Critical need for new definitions of “forest” and “forest degradation” in global climate change agreements. Conserv. Lett. 2, 226–232. https://doi.org/10.1111/j.1755-263X.2009.00067.x.

Scharf, L.L., 1991. Statistical Signal Processing: Detection Estimation and Time Series Analysis. Addison-Wesley.

Schwartz, M., 1984. Information Transmission, Modulation and Noise: A Unified Approach to Communication Systems. McGraw-Hill.

Shimabukuro, Y.E., Beuchle, R., Grecchi, R.C., Ached, F., 2014. Assessment of forest degradation in Brazilian Amazon due to selective logging and fires using time series of fraction images derived from Landsat ETM+ images. Remote Sens. Lett. 5, 773–782. https://doi.org/10.1080/2150704X.2014.967880.

Simula, M., 2009. Towards Defining Forest Degradation: Comparative Analysis of Existing Definitions. Forest Resources Assessment Programme Working Paper. Rome, Italy.

Souza, C., Barreto, P., 2000. An alternative approach for detecting and monitoring selectively logged forests in the Amazon. Int. J. Remote Sens. 21, 173–179. https://doi.org/10.1080/0143164021000101620.

Souza Jr., C., Siqueira, J., Sales, M., Fonseca, A., Ribeiro, J., Numata, I., Cochrane, M., Barber, C., Roberts, D., Barlow, J., 2013. Ten-year Landsat classification of deforestation and forest degradation in the Brazilian Amazon. Remote Sens. 5, 5493–5513. https://doi.org/10.3390/rs5115493.

Souza, C.J., Roberts, D., 2005. Mapping forest degradation in the Amazon region with Honos images. Int. J. Remote Sens. 26, 425–429. https://doi.org/10.1080/01431640500225901.

Souza, C.J., Roberts, D., Cochrane, M., 2005. Combining spectral and spatial information to map canopy damage from selective logging and forest fires. Remote Sens. Environ. 98, 329-343. https://doi.org/10.1016/j.rse.2005.07.013.

Thompson, I.D., Guariguata, M.R., Okabe, K., Bahamondez, C., Nasi, R., Heymell, V., Sabogal, C., 2013. An operational framework for defining and monitoring forest degradation. Ecol. Soc. 18, 20. https://doi.org/10.5751/ES-05443-180220.

Tuia, D., Volpi, M., Copi, L., Kanevski, M., Munoz-Mari, J., 2011. A survey of active learning algorithms for supervised remote sensing image classification. IEEE J. Sel. Top. Signal Process. 5, 606-617. https://doi.org/10.1109/JSTSP.2011.2139152.