

Evaluating remote sensing datasets and machine learning algorithms for mapping plantations and successional forests in Phnom Kulen National Park of Cambodia

Minerva Singh, Damian Evans, Jean-Baptiste Chevance, Boun Suy Tan, Nicholas Wiggins, Leaksmy Kong, Sakada Sakhoeun, Boun Suy Tan

▶ **To cite this version:**

Minerva Singh, Damian Evans, Jean-Baptiste Chevance, Boun Suy Tan, Nicholas Wiggins, et al.. Evaluating remote sensing datasets and machine learning algorithms for mapping plantations and successional forests in Phnom Kulen National Park of Cambodia. PeerJ, 2019, 7, pp.e7841. 10.7717/peerj.7841 . hal-02324835

HAL Id: hal-02324835
https://hal.science/hal-02324835

Submitted on 22 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluating remote sensing datasets and machine learning algorithms for mapping plantations and successional forests in Phnom Kulen National Park of Cambodia

Minerva Singh[1], Damian Evans[2], Jean-Baptiste Chevance[3], Boun Suy Tan[4], Nicholas Wiggins[5], Leaksmy Kong[2] and Sakada Sakhoeun[3]

[1] Imperial College, Centre of Environmental Policy, London, United Kingdom
[2] École Française d'Extrême-Orient, Paris, France
[3] Phnom Kulen Program, Archaeology and Development Foundation, London, United Kingdom
[4] Angkor International Research and Documentation Centre, Siem Reap, Cambodia, Siem Reap, Cambodia
[5] School of Earth and Environmental Sciences, University of Queensland, St Lucia, Australia

ABSTRACT

This study develops a modelling framework by utilizing multi-sensor imagery for classifying different forest and land use types in the Phnom Kulen National Park (PKNP) in Cambodia. Three remote sensing datasets (Landsat optical data, ALOS L-band data and LiDAR derived Canopy Height Model (CHM)) were used in conjunction with three different machine learning (ML) regression techniques (Support Vector Machines (SVM), Random Forests (RF) and Artificial Neural Networks (ANN)). These ML methods were implemented on (a) Landsat spectral data, (b) Landsat spectral band & ALOS backscatter data, and (c) Landsat spectral band, ALOS backscatter data, & LiDAR CHM data. The Landsat-ALOS combination produced more accurate classification results (95% overall accuracy with SVM) compared to Landsat-only bands for all ML models. Inclusion of LiDAR CHM (which is a proxy for vertical canopy heights) improved the overall accuracy to 98%. The research establishes that majority of PKNP is dominated by cashew plantations and the nearly intact forests are concentrated in the more inaccessible parts of the park. The findings demonstrate how different RS datasets can be used in conjunction with different ML models to map forests that had undergone varying levels of degradation and plantations.

Subjects Biodiversity, Ecosystem Science, Data Science, Forestry, Spatial and Geographic Information Science
Keywords Tropical forests, Deforestation, Plantations, SE Asia, Remote sensing, Landsat, ALOS PALSAR, Machine learning, LiDAR, Support vector machines

Submitted 5 March 2019
Accepted 5 September 2019
Published 22 October 2019

Corresponding author
Minerva Singh,
minerva_singh@yahoo.co.in

Academic editor
Danlin Yu

Additional Information and Declarations can be found on page 15

DOI **10.7717/peerj.7841**

Copyright
2019 Singh et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

INTRODUCTION

Conversion of forests to plantations for commodity production is a leading cause of permanent forest loss globally. Analysis of satellite imagery from 2000–2015 revealed that approximately 27% of global forest loss could be attributed to plantation creation (*Curtis et al., 2018*). The forests of tropical Asia, and especially lowland forests of insular

Southeast Asia, are the worst affected by the global demand for commodities (such as oil palm and rubber) and the subsequent conversion of forests to agricultural plantations (*Koh & Wilcove, 2008*; *Edwards et al., 2010*; *Edwards et al., 2013*). Commercial agricultural plantations such as rubber and acacia are also encroaching in the forests of continental Asia (*Torbick et al., 2017*; *Zeng et al., 2018*). Plantations for commodity crops now exist in Laos, China, Vietnam and Myanmar (*Phompila et al., 2014*; *Grogan et al., 2015*; *Chen et al., 2016*).

In addition to plantations, many tropical Asian ecosystems are significantly dominated by successional forests (*Singh et al., 2014*); forests that regrow after forest clearance or deforestation (*Lu et al., 2014*). While the role of primary tropical forests in preserving biodiversity is well-established, logged and degraded forests too can provide a valuable reservoir for endangered biodiversity and carbon stocks (*Woodcock et al., 2011*; *Edwards & Laurance, 2013*; *Mukul, Herbohn & Firn, 2016*; *Lennox et al., 2018*). Thus, the retention of regenerating and logged forests has been recommended as a means of preventing a biodiversity collapse in SE Asia (*Wilcove et al., 2013*). However, the ability of successional forests to provide habitat for species and store carbon depends on the age, structure and level of degradation of the forest (*Edwards et al., 2011*; *Singh et al., 2014*). Mapping the spatial extent of plantations and differentiating between forests of different successional ages and ecological conditions is important for the conservation of Asia's tropical forests.

Remote sensing (RS) data have been extensively used for mapping tropical ecosystems. Landsat data which can go back to the mid-1970s with spatial resolution of 30-m have been used extensively to map land use types and changes in the tropics (*Grogan et al., 2015*; *Maryantica & Lin, 2017*; *Midekisa et al., 2017*). Landsat data have proven useful in mapping and monitoring forest disturbance and recovery at global and continental scales (*Gong et al., 2013*; *Midekisa et al., 2017*; *DeVries et al., 2015*; *Liu et al., 2017a*). Single-date and multi-temporal Landsat data have helped map accurately the extent of plantations and successional forests in the tropics (*Connette et al., 2016*; *Kelley, Pitcher & Bacon, 2018*). Conventional approaches for using remote sensing data require users to download and process several terabytes of data. This is time-consuming and often out-of-the-reach for many practitioners. Google Earth Engine (GEE) is a cloud-based platform which allows the user to process RS data, including Landsat data using inbuilt pre-processing algorithms (*Gorelick et al., 2017*). Landsat data obtained via the GEE platform have been used to quantify land cover types and changes at different scales and in different ecosystems (*Midekisa et al., 2017*; *Tsai et al., 2018*).

Over the past few years, Landsat data were also used in conjunction with machine learning (ML) algorithms to distinguish between and map different land use types, including natural forests, anthropogenic land use types (including plantations) and degraded forests (*Sesnie et al., 2010*; *Connette et al., 2016*). ML algorithms were also implemented on GEE obtained Landsat data to map forest types, examine spectral properties and delineate shade-grown coffee in Nicaragua (*Kelley, Pitcher & Bacon, 2018*) and the oil palm plantations in Indonesia (*Lee et al., 2016*). Some of the most common machine learning techniques for carrying out land cover classification and mapping include, support vector machines (SVM), random forests (RF) and artificial neural networks (ANN). These have been

implemented in a variety of different settings, including mapping heterogenous tropical landscapes (*Sesnie et al., 2010*; *Szuster, Chen & Borger, 2011*; *Gong et al., 2013*; *Paneque-Gálvez et al., 2013*; *Shiraishi et al., 2014*; *Chuang & Shiu, 2016*; *Thanh Noi & Kappas, 2018*; *Hartling et al., 2019*). The ability of these common ML algorithms to distinguish between and map forests that have undergone varying levels of degradation and plantations in the human-modified forests of the Greater Mekong region countries needs to be evaluated further for the different forest ecosystems of the region in order to support conservation management.

The humid tropics suffer from the problem of persistent cloud cover which impedes the efficacy of optical RS data sources such as Landsat (*Asner, 2001*). This makes synthetic aperture radar (SAR) data increasingly being used in conjunction with optical data for land cover mapping in the tropics since it is relatively unaffected by cloud cover. Moreover, SAR data offer an additional benefit of having the ability to retrieve forest structure information (*Mitchard et al., 2012*). While it is possible to use Landsat and SAR data separately for land cover mapping, researches have shown that combining these datasets produce more accurate land cover maps in tropical ecosystems (*Lu et al., 2011*; *Lu et al., 2014*). LiDAR derived data also provides valuable information about the forest vertical structure that are useful in distinguishing different successional stages of regenerating tropical forests (*Marselis et al., 2018*). LiDAR data have been used with both Landsat and ALOS data to identify varying levels of forest degradation in Laos (*Singh et al., 2017a*) and for distinguishing between natural forests and cashew plantations in Cambodia (*Singh et al., 2018*).

In this study, we evaluate the ability of some common RS data and ML algorithms tools in mapping human modified forest-plantation tropical ecosystems, focusing on the cashew plantations that have encroached in an IUCN Category II protected forests of Phnom Kulen National Park (PKNP) in Cambodia. Despite its protected status, PKNP has experienced high deforestation and degradation rates. As with other protected areas in Cambodia, faces significant threats from local resource extraction activities (*Singh et al., 2018*). This research seeks to: (a) examine the efficacy of different RS data sources: Landsat, ALOS and LiDAR (either alone or in conjunction with each other) to delineate the different land cover types; (b) compare the performance of three commonly used for land cover classifications ML algorithms (SVM, RF and ANN) in delineating and mapping forests that have undergone varying levels of degradation and cashew plantations; and (c) assess the structural and spectral properties of the different land use types in the area.

MATERIALS & METHODS

Study Area

PKNP, mainly dominated by semi-evergreen forests, is located in Siem Reap Province in north-western Cambodia (Fig. 1). At 37,380 hectares, it covers the entire Kulen plateau which is a critical area for biodiversity and a significant component of the regional watershed. It is comprised mostly of regenerating and secondary forests along with severely

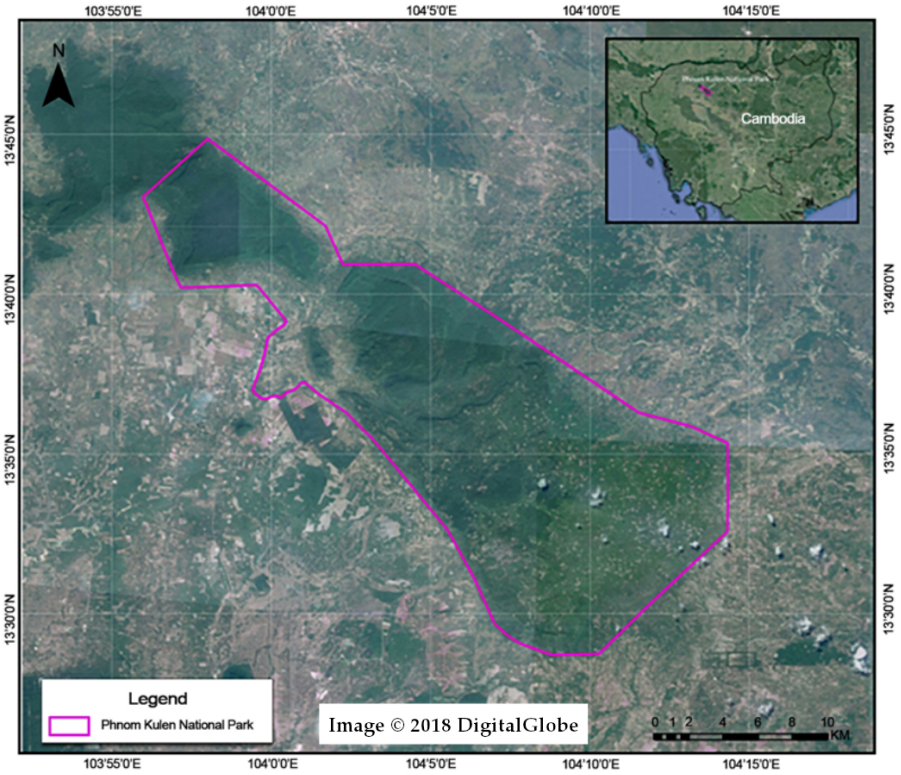


Figure 1 **Location of the Phnom Kulen National Park (PKNP) in Cambodia.**
Full-size DOI: 10.7717/peerj.7841/fig-1

degraded forests, cassava and cashew plantations, mostly located in the southern part (*Singh et al., 2018*).

Table 1 shows seven main land use classes identified in the area (*Singh et al., 2017a*; *Sirro et al., 2018*). The data pertaining to the geo-location of these land use types were collected in the field using a handheld GPS. These geo-locations were then verified using high-resolution imagery available on Google Earth.

Remote sensing data

Landsat 8 data

The Landsat 8 Operational Land Imager TOA (top-of the-atmosphere) data (30-m spatial resolution) were obtained and processed using the Google Earth Engine (GEE) platform (*Dong et al., 2016*; *Kelley, Pitcher & Bacon, 2018*). The GEE derived Landsat Annual TOA Reflectance Composite data was created by taking the median value from image bands from the target year, plus or minus one year (*Midekisa et al., 2017*). TOA corrected Landsat data have been extensively used for land cover mapping and classification in different tropical ecosystems, including tropical Asia (*Ramdani, Moffiet & Hino, 2014*; *Johnson & Iizuka, 2016*; *Dong et al., 2015*; *Pimple et al., 2018*).

Table 1 Land use/cover types and their characteristics.

Land use/cover type	General characteristics
Cashew Plantations	Cashew monocultures (created as result of clear felling forests).
Lightly Logged Forests	There was no previous record of these forests having undergone extensive selective logging and/or resource extraction. Qualitative field monitoring by JBC indicates light selective logging may have started occurring after 2015.
Degraded Forests	These forests have undergone several rounds of logging and resource extraction. These are mostly scrub forests with less than 50% canopy cover.
Selectively Logged Forests	These forests had a greater than 50% forest cover but most of the large Dipterocarp and rosewood trees had been selectively logged. Selective logging of these species was ascertained by the presence of tree stumps and confirmed by the forest rangers.
Regenerating for More Than 10 Years	These forests were clear felled/burnt for agricultural use more than a decade ago and were regenerating for more than 10 years as of March 2016.
Regenerating for Less Than 10 Years	These forests were clear felled/burnt for agricultural use less than a decade ago and were regenerating for less than 10 years as of March 2016.
Bare Earth/New Cassava Plantations	Cassava plantations (as of 2016) which had the crops, scrubby vegetation and bare earth patches.

ALOS L-band data

L-Band ALOS PALSAR data (25-m spatial resolution) were obtained in 2016 from Japan Aerospace Exploration Agency. These data have dual polarization HH (horizontal transmit, horizontal receive) and HV (horizontal transmit, vertical receive). These data are provided as digital numbers (DN). In the backscatter values (σ^0) (Eq. (1)), the DN values of both HH and HV were converted to the normalized radar cross section (*Avtar et al., 2013*; *Singh et al., 2017a*):

$$\sigma 0 = 10 \bullet log 10(DN)^2 + CF. \quad (1)$$

An enhanced Lee filter was implemented using a 3x3 kernel size to reduce the noise caused by speckles (*Morel et al., 2011*). These backscatter values are strongly associated with forest structural properties, including canopy cover (*Mitchard et al., 2011*; *Deus, 2016*; *Singh et al., 2017a*). The backscatter HH and HV values were then used to compute two radar-based indices- Radar Forest Degradation Index (RFDI) and the ratio of HH and HV (HH/HV) (Eq. (2)).

$$RFDI = (HH - HV)/(HH + HV), \quad (2)$$

This index varies from 0 to 1, with values closer to 1 indicating a more open canopy and thus higher levels of degradation (*Singh et al., 2017a*; *Singh et al., 2018*). This index was used to map different forest cover types in Gabon (*Mitchard et al., 2012*) and to distinguish forest patches that had undergone varying levels of degradation in Laos (*Singh et al.,*

2017a). This index was also used to differentiate and map cashew plantations and forests in a protected area in Cambodia (Singh et al., 2018) and for land cover mapping in northern Tanzania (Deus, 2016). The HH/HV ratio has been used for land cover classification (Deus, 2016; Ali, Qazi & Aslam, 2018) and was proven useful in differentiating between forest and non-forested areas in the Riau province in Indonesia (Quegan et al., 2014).

LiDAR

The LiDAR data were collected over the study area in 2016 using a Leica ALS60 laser system and a 40 megapixel Leica RCD105 medium-format camera in an external pod mounted on the left skid of a Eurocopter AS350 B2 helicopter. These data have a density of 15 points/m^2. To achieve requisite accuracy and point density, flights were at altitudes of 800–1,000 m above-ground level at a speed of 80 knots, with the ALS70 configured to Multipulse in Air (MPiA). The pulse rate was 500 kHz with a scan angle of 45° from nadir and a swath side-lap of 50% (i.e., almost all terrain was scanned twice from different angles). Aircraft attitude was measured by a Honeywell CUS6 IMU at a rate of 200 kHz and positional data was logged at 2 Hz using a survey-grade L1/L2 GNSS receiver mounted in the tail rotor assembly (Evans, 2016). The processing of these data to obtain a canopy height model (CHM) at 25m spatial resolution was carried out as done by Singh et al. (2018). CHMs provide a 3D representation of tree canopy heights. These were obtained by deriving a digital elevation model (DEM) from ground LiDAR returns and the vegetation/non-ground returns were used to a digital surface model (DSM). These were subtracted from each other to obtain a CHM (Singh et al., 2015). CHMs have been extensively used for tree species identification and biomass mapping and changes in Cambodia (Singh et al., 2015; Singh et al., 2016; Singh et al., 2018).

Statistical data analysis

Spectral and structural properties evaluation for the different land use types

Prior to any statistical analysis, the Landsat data were resampled to 25-m to be at the same spatial resolution as ALOS PALSAR and LiDAR-derived data. All the Landsat bands were used in the analysis to enable better discrimination between the different land cover classes (Bhagwat et al., 2017). Reflectance values for the Landsat bands, ALOS-derived variables and LiDAR CHM were extracted for the geolocations of the different land use types collected in the field. Kruskal-Wallis (KW) tests were applied using the using the R statistical software package, version 3.3.2 (R Core Team, 2013) to examine whether the Landsat based spectral and ALOS and LiDAR-derived forest structure variables (HH, HV, HH/HV, RFDI and LiDAR CHM) varied across the different land cover classes, including cashew plantations. This is a nonparametric test that does not need the assumption of normally distributed data (Gaveau et al., 2009; Singh et al., 2017a). The "pgirmess" package of R (Giraudoux, 2018) was used to carry out post hoc analysis to identify which land use types were significantly different in terms of their spectral and structural properties as done by Rembold et al. (2017) and Singh et al. (2017a).

Land cover mapping

Three commonly used ML algorithms for tropical land cover classification, SVM, RF, and ANN were employed for classification purposes. The training used 70% of the geolocation data and the other 30% for testing the models (*Rogan et al., 2008*; *Wang et al., 2015*). Classification metrics including overall accuracy, kappa coefficients, user's and producer's accuracy were computed from the testing data. Overall accuracy computes how classified pixels compare with the field collected test data. Producer's accuracy measures how well the real world classes can be classified. User's accuracy quantifies the likelihood of a classified pixel matching the land cover type of its corresponding real-world location (*Rwanga & Ndambuki, 2017*). Additionally, sensitivity and specificity were calculated using Eq. (3) to evaluate the ability of the different ML and remote sensing data combinations to discriminate between specific land cover classes (*Silva et al., 2017*), in this case, cashew plantations from other land use types, based on their geometric mean (G).

$$G = \sqrt{(sensitivity * specificity)}, \tag{3}$$

Classification errors can result in the area and extent of a class to be overestimated or underestimated (*Alatorre et al., 2011*). Sensitivity is the proportion of true positives classified by the algorithm and specificity is the proportion of true negatives. In many cases, high sensitivity is associated with low specificity and vice versa. Computing the geometric mean allows us to combine the two and balance between detecting and not detecting a certain land cover class (*Silva et al., 2017*; *Silva, Bacao & Caetano, 2017*).

Support vector machine

SVM is a supervised nonparametric statistical learning method. The SVM aims to find a hyperplane that optimally separates linearly-separable classes (*Singh et al., 2015*). In the simplest form, SVMs are binary classifiers that assign the given test sample to one of two possible classes. The SVM algorithm is extended to non-linearly separable classes by mapping samples in the feature space to a higher dimensional feature space using a kernel function. Data projection via hyperplanes allows the algorithm to discriminate between the different land use classes (*Kanniah et al., 2015*). SVMs are particularly appealing in remote sensing due to their ability to successfully handle small training datasets, often producing higher classification accuracy than traditional methods (*Maryantica & Lin, 2017*). It is also a non-parametric algorithm, i.e., it is not influenced by the underlying data distribution (*Tiwari et al., 2016*).

The SVM algorithm was implemented using the *caret* package of the R programming language (*Kuhn, 2015*). The SVM algorithm in the "caret package" is implemented via the kernlab package of the R programming language (*Karatzoglou et al., 2004*). The radial bias function (RBF) kernel which has been extensively used for land cover applications (*Thanh Noi & Kappas, 2018*) was used for this research. RBF-kernel based SVM has produced robust results with complex land cover related datasets (*Ghosh & Joshi, 2014*) and use of SAR data for tropical forest mapping in Indonesia (*Trisasongko et al., 2017*). The use of the radial bias kernel function requires tuning of two parameters: cost of constraints violation (C) and gamma. The former accounts for the over-fitting of the model while the latter

Table 2 Parameters of the classifiers used in building the classification models.

Classifier	Parameter		
ANN	size	decay	rang
	10	5e-6	0.1
RF	mtry	ntree	
	6	500	
SVM	kernel	cost	gamma
	"radial"	10	0.1
	"optimal"	9	1

controls the shape of the hyperplane. A cost function value (C) of 10 and a gamma value of 0.1 were finally used as tuning parameters (*Wang et al., 2015*; *Chuang & Shiu, 2016*). A 10-fold cross-validation repeated three times for the model building (*Ghosh & Joshi, 2014*). The SVM tuning parameters are tabulated in Table 2.

Random forests

RF is a decision tree-based ML algorithm. Single decision tree models are built through recursive partitioning, wherein the response variable is iteratively divided into groups sequentially with group 'purity' increasing with each division. RF models fit multiple decision trees to input data using a random subset of the input variables for each tree constructed. An average derived from the multitude of trees is used to form the predictive model (*Singh et al., 2015*). RF models are non-parametric, handle a large number of correlated input variables and skewed data and prevent overfitting. This makes RF models very useful for mapping heterogeneous landscapes (*Attarchi & Gloaguen, 2014*). For classification purposes, an ensemble of individual decision- tree classifiers are created and these are combined using a majority voting scheme (*Singh et al., 2018*). The RF classifier of this study was implemented using the "caret" package of the R programming language (*Kuhn et al., 2016*). Caret package implements the RF algorithm using the "randomForest" package (*Liaw & Wiener, 2002*). Since classification accuracy is sensitive to tuning parameters, the number of trees (n) was fixed at a default value of 500 (*Li et al., 2016*; *Raczko & Zagajewski, 2017*). The other tuning parameters were identified using the in-built *tuneGrid* function of the algorithm. The *tuneGrid* is a function that searches for optimal mtry values (the number of random predictors used for splitting at a node) given the data (*Ghosh & Joshi, 2014*). This parameter tuning process was performed following a 3 time repeated 10-fold cross-validation process (*Chuang & Shiu, 2016*; *Ghosh, Sharma & Joshi, 2014*). The RF tuning parameters are tabulated in Table 2.

Artificial neural networks

ANN is a biologically inspired ML algorithm that has been extensively used for building predictive regression models with nonlinear ecological data (*Foody & Cutler, 2006*). Multi-Layer Perceptrons (MLPs) are the most common type of ANN used for remote sensing studies and were used for this study. MLP networks use a back propagation learning principle and one hidden layer, and are useful for modeling non-linear relationships

(*Jensen, Fang & Minhe, 1999*). This setup is typically composed of an input layer (for feeding in predictors), an output layer (a response variable) and one hidden layer which accounts for the non-linearities in the data (*Liu et al., 2017b*). The model is trained by predicting the response from patterns learned from a training data set (*Chambers et al., 2016*). By comparing the current output layer to the desired output response, the difference between the two can be obtained and used to adjust weights within the network. The goal is to find weights that produce results that closely resemble the target response. This iterative modelling process is repeated until the predicted response values meet a given level of accuracy (*Ingram, Dawson & Whittaker, 2005*). The MLP based ANN was implemented using the remote sensing data as predictors through the *nnet*'' package of R (*Venables & Ripley, 1997*). The nnet algorithm simulates feed-forward neural networks with a single hidden layer with a backpropagation algorithm as the training algorithm (*Werbos, 1994*). Implementation of the algorithm needs the tuning of three important parameters, size, decay and maxit. The size parameter sets the number of units in the hidden layer and needs to be tuned. The decay parameter controls the weight decay, and maxit sets the maximum number of iterations; they are both left at their default values (*Li et al., 2016*; *Raczko & Zagajewski, 2017*). A logistic activate/transfer function and the quasi-Newton optimization algorithm that does not use the parameters, such as learning rate and momentum, were used (*Venables & Ripley, 1997*). The tuning parameters are tabulated in Table 2.

RESULTS

Spectral and structural properties of the different land use types

The reflectance curves of different land use classes across different Landsat bands are shown in Fig. 2. It indicates that cashew plantations had the highest reflectance (out of all land cover types) in band 4. The reflectance curves of different natural forests were very close to each other, especially in band 3, 4 and 5.

KW tests revealed that in band 3, cashew plantations were significantly different from all the land cover classes, except cassava and selective logging (KW = 113.8, $p < 0.05$). The degraded forests were significantly different from the virtually intact forests only. The near-intact forests, regenerating forests and selectively logged forests were not significantly different from each other in band 3. In band 4 and 5, the cashew plantations were significantly different from all land cover classes, except the cassava plantations (KW = 68.3, $p < 0.001$). The natural forest types: degraded forests, regenerating forests (both more than and less than 10 years), near-intact and selectively logged forests were not significantly different from each other in band 4. In band 5, the degraded forests varied significantly from all natural forest types, except the selectively logged forests.

KW test also revealed that the HH backscatter values did not vary significantly across the different land use classes. However, as Fig. 3 shows, the other structural metrics- backscatter HV, HH/HV ratio, RFDI and LiDAR-derived CHM varied significantly across the different land use classes (KW = 40.58, 35.2, 35.7, 109.9, respectively, $p < 0.005$).

Post-hoc analysis revealed that the HV backscatter varied significantly between both cashew plantations and near intact forests and between cashew plantations and regenerating

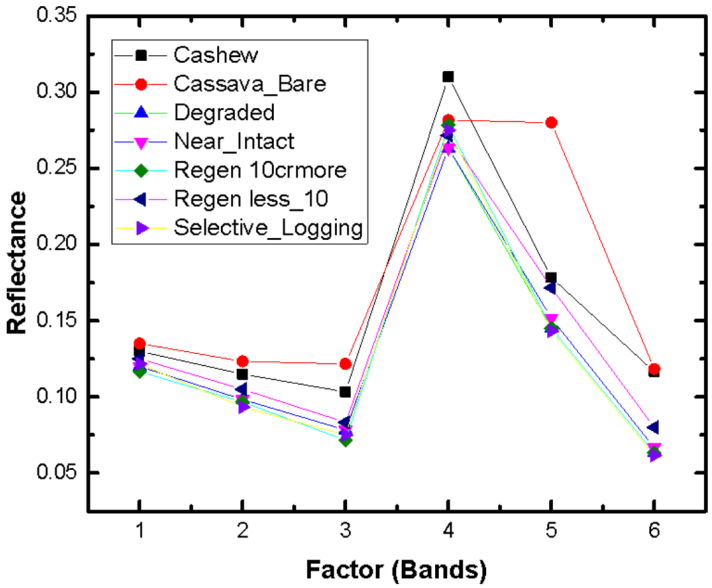


Figure 2 Spectral Behaviour of the Different Landcover Classes in PKNP.

Full-size DOI: 10.7717/peerj.7841/fig-2

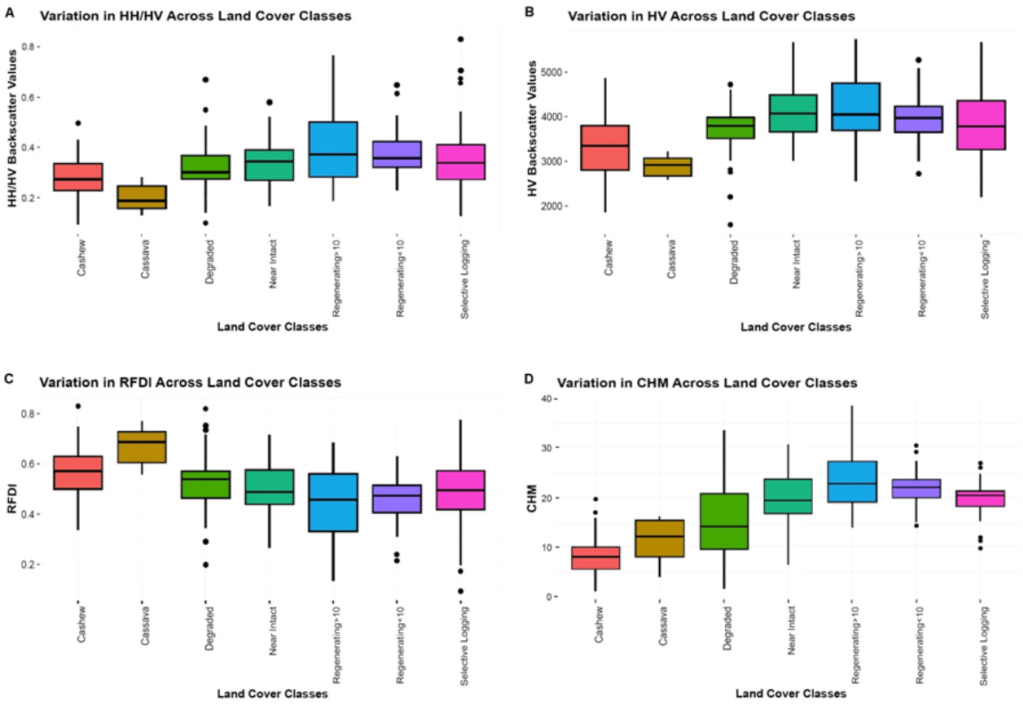


Figure 3 (A) Varaition in HH/HV across different land cover classes, (B) variation in HV across the different landcover classes, (C) variation in RFDI across different land cover classes, (D) variation in CHM across different land cover classes.

Full-size DOI: 10.7717/peerj.7841/fig-3

Table 3 Overall accuracy of the different ML models and their corresponding Kappa values.

Different RS datasets used	SVM	RF	ANN
Landsat spectral bands	56.0% (0.45)	62.0% (0.54)	28.5% (0.16)
Landsat spectral bands + ALOS metrics	95.0% (0.94)	63.0% (0.54)	33.0% (0.21)
Landsat spectral bands + ALOS metrics LiDAR CHM	98.0% (0.98)	66.3% (0.59)	31.0% (0.19)

forests (which had been regenerating for both more and less than ten years). The HH/HV ratio and RFDI varied significantly between the cashew plantations and regenerating forests. These metrics varied significantly between cassava plantations, near intact forests, regenerating forests and selectively logged forests. Cashew plantations had a mean RFDI value of 0.57 while the regenerating forests had average values of 0.43 and 0.45 respectively for the forests that had been regenerating for more than 10 years and less than 10 years. LiDAR-derived heights (from the CHM) were significantly different for cashew plantations and the other land cover types in question. Out of all the land cover types, cashew plantations had the lowest average values (mean = 8.4, sd = 3.9) while the average vertical canopy heights of the other classes varied from 11 m–23 m.

Comparison of classification results with different datasets and ML techniques

Of all the RS datasets combinations considered, the Landsat-only band data performed the worst with all ML algorithms. The overall accuracy of the kernel based SVM model was 56.0% when all the spectral bands were considered. With RF and ANN, the overall classification accuracy was 62.0% and 28.5% respectively (Table 3). The kappa coefficients of the SVM, RF and ANN models were 0.45, 0.54 and 0.16 respectively. However, the RF model also produced a geometric mean of 0.9 for the cashew plantations, indicating a strong ability to discriminate cashew plantations. In addition to the Landsat spectral bands, the four ALOS derived products were included and the three ML techniques were implemented on these. The overall accuracies obtained with SVM, RF and ANN were 95.0%, 63.0% and 33.0% respectively. The kappa coefficients of the SVM, RF and ANN models were 0.94, 0.54 and 0.21 respectively. The SVM model had a G-value of 1 for cashew plantations while the RF model produced a G-value of 0.82. The SVM variable importance analysis revealed that out of all the ALOS-PALSAR variables included in the mapping process, the HV band and RFDI were most useful in discriminating between the different land use classes.

The highest overall accuracy was obtained when LiDAR CHM data were included along with the Landsat and ALOS data bands. The overall accuracies obtained with SVM, RF and ANN were 98.0%, 66.3% and 31.0% respectively. The kappa coefficients of the SVM, RF and ANN models were 0.98, 0.59 and 0.19 respectively. The SVM model had a G-value of 1 for cashew plantations while the RF model produced a G-value of 0.72. The SVM variable importance analysis revealed that LiDAR CHM was more important than the ALOS-PALSAR and Landsat variables included in the mapping process.

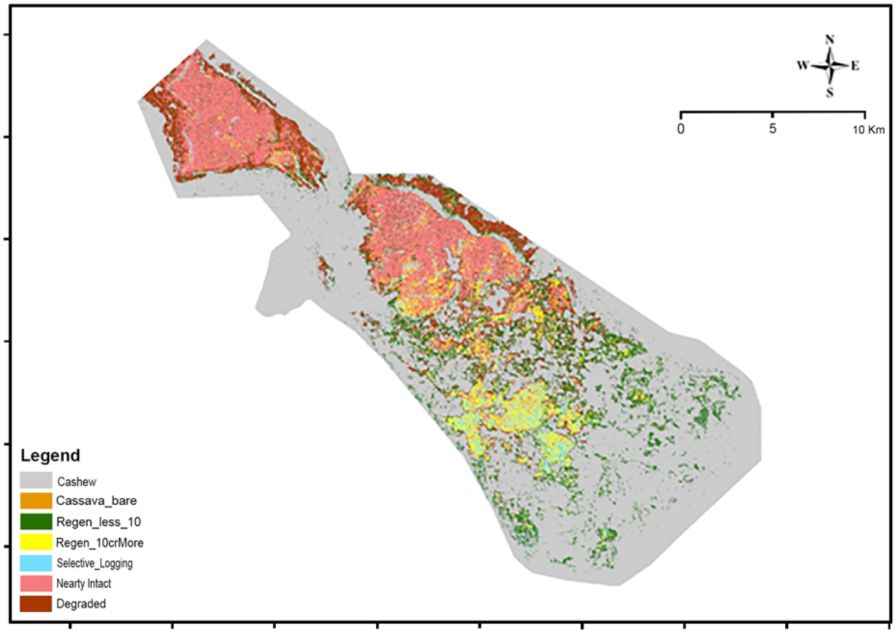


Figure 4 Land Cover Map of PKNP.

Full-size DOI: 10.7717/peerj.7841/fig-4

Table 4 Comparison of producer's and user's accuracy levels for Landsat-ALOS-LiDAR data SVM model.

	Cashew	Bare Earth	Regenerating for less than 10 years	Regenerating for more than 10 years	Selective logging	Lightly logged forests	Degraded Forests
% User's accuracies	100	100	100	100	100	100	87.5
% Producer accuracies	100	100	100	94	100	100	100

Land cover map

A combination of Landsat-ALOS data within a framework of an SVM classification algorithm had an overall accuracy of 95.0%. However, the inclusion of LiDAR-derived canopy heights improved the overall classification accuracy to 98.0%. This model was thus used for producing a land cover map of PKNP (Fig. 4). The producer and user accuracies of the SVM model which utilized the Landsat, ALOS and LiDAR data have been tabulated in Table 4. All the land use classes were identified accurately. The analysis also revealed that more than 60.0% of PKNP is dominated by cashew plantations. Nearly intact forests which cover 13.3% are clustered in the northern part of PKNP, which is relatively inaccessible.

DISCUSSION

Landsat spectral data models have been used for both mapping land cover and for identifying land cover change across different tropical forest ecosystems (*Connette et al., 2016*; *Hansen et al., 2016*; *Bhagwat et al., 2017*; *Hurni et al., 2017*; *Shen, Li & Wei, 2017*). The ability of these data to distinguish between different land cover/forest types varies considerably across different tropical ecosystems. In this study revealed that the separability

between the different land cover classes (especially between the different natural forests and cashew plantations) is limited across the Landsat bands. The Landsat spectral band model utilized for land cover mapping had an overall accuracy of 68.0%. This is similar to the overall accuracy of a Landsat spectral data only map developed using an ML model developed to map 32 different land use classes in Costa Rica (*Sesnie et al., 2008*). The relatively low overall classification accuracy of the Landsat model, when used in isolation, can be attributed to low spectral separability of different tropical forest/land cover types in Landsat imagery (*Tottrup, 2004*). This research is the first of its kind which has identified low spectral separability exists between natural forest types, and between cashew plantations and natural forests using Landsat-only models. This is cognizant with the existing body of literature. The phenomenon of lower spectral separability between plantation monocultures and natural forests has been observed in other plantation types such as rubber (*Li & Fox, 2012*), oil palm plantations (*Broich et al., 2011*) and teak (*Win, Reiji & Shinya, 2009*).

To mitigate this weakness, in several applications, ALOS L-band data and Landsat spectral bands have been used jointly to map different tropical forest types (*De Alban et al., 2018*) including plantation monocultures (*Kou et al., 2015*; *Chen et al., 2018*). HH and HV backscatter values are known to be influenced by the structure of monocultures such as oil palm and rubber plantations (*Torbick et al., 2017*). This research too discovered that inclusion of backscatter derived metrics helps improve the classification accuracy and distinguish between the forests that had undergone varying levels of degradation and cashew plantations. The inclusion of ALOS derived metrics produced higher classification accuracy for mapping the successional forests of the Brazilian Amazon than Landsat-only model (*Li et al., 2011*). A combination of Landsat spectral bands and ALOS L-band derived metrics (including RFDI) were used within an SVM model to differentiate between different land use classes in northern Tanzania with a high level of accuracy (*Deus, 2016*). A combination of Landsat spectral bands and ALOS data previously outperformed individual sensor data models for mapping different land use types (including human-modified landscapes such as plantations) in southern Myanmar (*De Alban et al., 2018*). The research also establishes that the HH-band of the ALOS data did not vary appreciably between the different land cover classes. Previous research discovered that HV bands can distinguish between different tropical land cover types better than the HH backscatter values (*Dong et al., 2014*; *Singh et al., 2017b*). The HV backscatter was able to discriminate cashew plantations from intact and regenerating forests. While this is the first time these metrics have been used for mapping cashew plantations, previous research has established that HV backscatter values are sensitive to and can help differentiate oil palm plantations from natural forests in Cameroon (*Li et al., 2015*). HV backscatter was found to be sensitive to both oil palm and rubber plantations in Kalimantan and Myanmar respectively (*Torbick et al., 2017*). This, in turn, may be attributed to the fact that HV backscatter interacts closely with the vertical structure of the trees (*Mitchard et al., 2011*) which makes it sensitive to the vertical structural differences between plantation monocultures and natural forests. Derived products, notably the HH/HV ratio, have distinguished woody plantations such as rubber from other forests in SE Asia (*Win, Reiji & Shinya, 2009*; *Dong et al., 2014*). In this study too, the HH/HV ratio

of the cashew and cassava plantations was different from regenerating, selectively logged and intact natural forests. It was also discovered that in addition to the HV backscatter values, RFDI were important variables for mapping land cover types. RFDI was included as it can help differentiate between different forest/vegetation types in the tropics (*Mitchard et al., 2012*). This is the first time RFDI was used to map a plantation monoculture. This research establishes that RFDI values vary significantly between cashew plantations and natural forests. Previously, RFDI was used successfully to differentiate between forests that had undergone varying levels of degradation in Laos (*Singh et al., 2017a*), to distinguish forest types in northern Tanzania (*Deus, 2016*) and forests and non-forests (*Ningthoujam et al., 2016*).

In this research, the inclusion of structural attributes (derived from LiDAR and ALOS L-Band data) improved the classification accuracy of the land cover models considerably (when used in conjunction with the SVM algorithm). Out of all the data, the LiDAR-derived canopy height was most sensitive to the differences between the land cover classes. LiDAR derived heights too were found to be significantly different between cashew plantations and natural forests, indicating that the former are structurally different from natural forests. LiDAR derived heights were previously used to differentiate between natural forests and cashew plantations accurately in the PKNP (*Singh et al., 2018*) and forests that have undergone varying levels of degradation in southern Laos (*Singh et al., 2017a*). LiDAR-derived vertical height estimates also proved effective in differentiating between different types of tropical moist forests in the Neotropics (*Kennel et al., 2013*).

The model which contained both ALOS and LiDAR-derived metrics had the highest overall accuracy of 98.0% although the Landsat-ALOS model had an overall accuracy of 95.0% with SVM classification. The RF model displayed a robust ability to identify cashew plantations even with the Landsat-only data (although the overall accuracy was much lower in this case). Both SVM and RF are known to outperform other classifiers for mapping and differentiating between different tropical forest types (*Kennel et al., 2013*). SVM models, especially give better results than other classifiers (including ANN) when used with smaller training sets and multiple RS band data (*Pal & Mather, 2005*). Three supervised classification methods, Maximum Likelihood Classification (MLC), SVM and ANN were tested to carry out classification of an urban ecosystem in Mexico. In this study, SVMs classification method performed better than both ANN and MLC. Land cover mapping of a complex landscape was conducted using three machine learning algorithms (RF, SVM, and ANN). SVM produced average overall accuracy of 86.63% while ANN models produced average overall accuracy of 73.55% (*Li et al., 2016*). Both SVM and RF models gave better classification results with multiple RS data sources as compared to other classifiers, including ANN (*Xie et al., 2019*). Owing to their ability to outperform more traditional classifiers for land use mapping in the tropics, SVM models have been used for differentiating between forest and land cover types in a number of tropical Asian ecosystems (*Kanniah et al., 2015*; *Tiwari et al., 2016*). SVM based classifiers were used to map healthy mangroves and differentiate them from other land use types in the Bay of Bengal region (*Singh et al., 2014*) and for land cover mapping in Vietnam (*Thanh Noi & Kappas, 2018*).

CONCLUSIONS

The problem of cashew plantation encroachment in PKNP is a microcosm of the challenges faced by the protected areas of SE Asia. This study highlights the potential to harness multi-sensor data and different ML algorithms to distinguish between and map forests and plantations across a heterogeneous landscape. While LiDAR data can produce highly accurate classifications, owing to the expense, these data may be difficult to obtain for conservation managers in tropical Asia. This research used freely available Landsat and ALOS data and helped identify how these could be used to give an overall accuracy results comparable to the model that included the commercially available LiDAR data. Future research will undertake similar mapping techniques in other protected areas of tropical Asia to quantify the magnitude of encroachment by different plantations.

ACKNOWLEDGEMENTS

This research was authorised by the APSARA National Authority as part of a research partnership between the Cambodian government, the Archaeology and Development Foundation, the University of Sydney, and the French Institute of Asian Studies (EFEO). The authors are grateful to the eight partners of the Khmer Archaeological LiDAR Consortium for the provision of LiDAR data from 2012, and to the three partners of the Cambodian Archaeological LiDAR Initiative for the provision of LiDAR data for 2015. The authors are also thankful to the rangers and inhabitants of the Phnom Kulen National Park for their assistance in conducting this research.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This project is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 639828). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
European Union's Horizon 2020 research and innovation programme: 639828.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Minerva Singh conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Damian Evans performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

- Jean-Baptiste Chevance contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.
- Boun Suy Tan and Sakada Sakhoeun contributed reagents/materials/analysis tools, approved the final draft.
- Nicholas Wiggins performed the experiments, approved the final draft.
- Leaksmy Kong performed the experiments, contributed reagents/materials/analysis tools, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available at Github: https://github.com/Jojo666/PKNP-Data.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7841#supplemental-information.

REFERENCES

Alatorre LC, Sánchez-Andrés R, Cirujano S, Beguería S, Sánchez-Carrillo S. 2011. Identification of mangrove areas by remote sensing: the ROC curve technique applied to the northwestern mexico coastal zone using landsat imagery. *Remote Sensing* **3**:1568–1583 DOI 10.3390/rs3081568.

Ali MZ, Qazi W, Aslam N. 2018. A comparative study of ALOS-2 PALSAR and landsat-8 imagery for land cover classification using maximum likelihood classifier. *The Egyptian Journal of Remote Sensing and Space Science* **21**:S29–S35 DOI 10.1016/j.ejrs.2018.03.003.

Asner GP. 2001. Cloud cover in landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing* **22**:3855–3862 DOI 10.1080/01431160010006926.

Attarchi S, Gloaguen R. 2014. Classifying complex mountainous forests with L-Band SAR and landsat data integration: a comparison among different machine learning methods in the hyrcanian forest. *Remote Sensing* **6**:3624–3647 DOI 10.3390/rs6053624.

Avtar R, Suzuki R, Takeuchi W, Sawada H. 2013. PALSAR 50 m mosaic data based national level biomass estimation in Cambodia for implementation of REDD+ mechanism. *PLOS ONE* **8**:e74807 DOI 10.1371/journal.pone.0074807.

Bhagwat T, Hess A, Horning N, Khaing T, Thein ZM, Aung KM, Aung KH, Phyo P, Tun YL, Okay AH, Neil A. 2017. Losing a jewel-Rapid declines in Myanmar's intact forests from 2002-2014. *PLOS ONE* **12**:e0176364 DOI 10.1371/journal.pone.0176364.

Broich M, Hansen MC, Potapov P, Adusei B, Lindquist E, Stehman SV. 2011. Time-series analysis of multi-resolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. *International Journal of Applied Earth Observation and Geoinformation* **13**:277–291 DOI 10.1016/j.jag.2010.11.004.

Chambers ME, Fornwalt PJ, Malone SL, Battaglia MA. 2016. Patterns of conifer regeneration following high severity wildfire in ponderosa pine—dominated

forests of the Colorado Front Range. *Forest Ecology and Management* **378**:57–67 DOI 10.1016/j.foreco.2016.07.001.

Chen B, Li X, Xiao X, Zhao B, Dong J, Kou W, Qin Y, Yang C, Wu Z, Sun R, Lan G. 2016. Mapping tropical forests and deciduous rubber plantations in Hainan Island, China by integrating PALSAR 25-m and multi-temporal Landsat images. *International Journal of Applied Earth Observation and Geoinformation* **50**:117–130 DOI 10.1016/j.jag.2016.03.011.

Chen B, Xiao X, Ye H, Ma J, Doughty R, Li X, Zhao B, Wu Z, Sun R, Dong J, Qin Y. 2018. Mapping forest and their spatial–temporal changes from 2007 to 2015 in tropical hainan island by integrating ALOS/ALOS-2 L-Band SAR and landsat optical images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**:852–867 DOI 10.1109/JSTARS.2018.2795595.

Chuang YC, Shiu YS. 2016. A comparative analysis of machine learning with WorldView-2 pan-sharpened imagery for tea crop mapping. *Sensors* **16**:594 DOI 10.3390/s16050594.

Connette G, Oswald P, Songer M, Leimgruber P. 2016. Mapping distinct forest types improves overall forest identification based on multi-spectral landsat imagery for Myanmar's Tanintharyi Region. *Remote Sensing* **8(11)**:882 DOI 10.3390/rs8110882.

Curtis PG, Slay CM, Harris NL, Tyukavina L, Hansen MC. 2018. Classifying drivers of global forest loss. *Science* **361**:1108–1111 DOI 10.1126/science.aau3445.

De Alban JDT, Connette GM, Oswald P, Webb EL. 2018. Combined landsat and L-Band SAR data improves land cover classification and change detection in dynamic tropical landscapes. *Remote Sensing* **10**:306 DOI 10.3390/rs10020306.

Deus D. 2016. Integration of ALOS PALSAR and landsat data for land cover and forest mapping in northern tanzania. *Land* **5**:43 DOI 10.3390/land5040043.

DeVries B, Decuyper M, Verbesselt J, Zeileis A, Herold M, Joseph S. 2015. Tracking disturbance-regrowth dynamics in tropical forests using structural change detection and Landsat time series. *Remote Sensing of Environment* **169**:320–334 DOI 10.1016/j.rse.2015.08.020.

Dong J, Xiao X, Kou W, Qin Y, Zhang G, Li L, Jin C, Zhou Y, Wang J, Biradar C, Liu J. 2015. Tracking the dynamics of paddy rice planting area in 1986–2010 through time series Landsat images and phenology-based algorithms. *Remote Sensing of Environment* **160**:99–113 DOI 10.1016/j.rse.2015.01.004.

Dong J, Xiao X, Menarguez MA, Zhang G, Qin Y, Thau D, Biradar C, Moore III B. 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sensing of Environment* **185**:142–154 DOI 10.1016/j.rse.2016.02.016.

Dong J, Xiao X, Sheldon S, Biradar C, Zhang G, Duong ND, Harazika M, Wikantika K, Takeuhci W, Moore III B. 2014. A 50-m forest cover map in Southeast Asia from ALOS/PALSAR and its application on forest fragmentation assessment. *PLOS ONE* **9**:e85801 DOI 10.1371/journal.pone.0085801.

Edwards FA, Edwards DP, Hamer KC, Davies RG. 2013. Impacts of logging and conversion of rainforest to oil palm on the functional diversity of birds in Sundaland. *Ibis* **155**:313–326 DOI 10.1111/ibi.12027.

Edwards DP, Hodgson JA, Hamer KC, Mitchell SL, Ahmad AH, Cornell SJ, Wilcove DS. 2010. Wildlife-friendly oil palm plantations fail to protect biodiversity effectively. *Conservation Letters* **3**:236–242 DOI 10.1111/j.1755-263X.2010.00107.x.

Edwards DP, Larsen TH, Docherty TDS, Ansell FA, Hsu WW, Derhé MA, Hamer KC, Wilcove DS. 2011. Degraded lands worth protecting: the biological importance of Southeast Asia's repeatedly logged forests. *Proceedings of the Royal Society B: Biological Sciences* **278**:82–90 DOI 10.1098/rspb.2010.1062.

Edwards DP, Laurance WF. 2013. Biodiversity despite selective logging. *Science* **339**:646–647 DOI 10.1126/science.339.6120.646-b.

Evans D. 2016. Airborne laser scanning as a method for exploring long-term socio-ecological dynamics in Cambodia. *Journal of Archaeological Science* **74**:164–175 DOI 10.1016/j.jas.2016.05.009.

Foody GM, Cutler MEJ. 2006. Mapping the species richness and composition of tropical forests from remotely sensed data with neural networks. *Ecological Modelling* **195**:37–42 DOI 10.1016/j.ecolmodel.2005.11.007.

Gaveau DLA, Epting J, Lyne O, Linkie M, Kumara I, Kanninen M, Leader-Williams N. 2009. Evaluating whether protected areas reduce tropical deforestation in Sumatra. *Journal of Biogeography* **36**:2165–2175 DOI 10.1111/j.1365-2699.2009.02147.x.

Ghosh A, Joshi PK. 2014. A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation* **26**:298–311 DOI 10.1016/j.jag.2013.08.011.

Ghosh A, Sharma R, Joshi PK. 2014. Random forest classification of urban landscape using Landsat archive and ancillary data: combining seasonal maps with decision level fusion. *Applied Geography* **48**:31–41 DOI 10.1016/j.apgeog.2014.01.003.

Giraudoux P. 2018. pgirmess: Spatial Analysis and Data Mining for Field Ecologists. R package version 1.6.9. *Available at https://CRAN.R-project.org/package=pgirmess*.

Gong P, Wang J, Yu L, Zhao Y, Zhao Y, Liang L, Niu Z, Huang X, Fu H, Liu S, Li C. 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing* **34**:2607–2654 DOI 10.1080/01431161.2012.748992.

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* **202**:16–27 DOI 10.1016/j.rse.2017.06.031.

Grogan K, Pflugmacher D, Hostert P, Kennedy R, Fensholt R. 2015. Cross-border forest disturbance and the role of natural rubber in mainland Southeast Asia using annual Landsat time series. *Remote Sens Environment* **169**:438–453 DOI 10.1016/j.rse.2015.03.001.

Hansen MC, Krylov A, Tyukavina A, Potapov PV, Turubanova S, Zutta B, Ifo S, Margono B, Stolle F, Moore R. 2016. Humid tropical forest disturbance alerts using

Landsat data. *Environmental Research Letters* **11**:034008 DOI 10.1088/1748-9326/11/3/034008.

Hartling S, Sagan V, Sidike P, Maimaitijiang M, Carron J. 2019. Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning. *Sensors* **19**:1284 DOI 10.3390/s19061284.

Hurni K, Schneider A, Heinimann A, Nong D, Fox J. 2017. Mapping the expansion of boom crops in mainland southeast asia using dense time stacks of landsat data. *Remote Sensing* **9**:320–346 DOI 10.3390/rs9040320.

Ingram JC, Dawson TP, Whittaker RJ. 2005. Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment* **94**:491–507 DOI 10.1016/j.rse.2004.12.001.

Jensen R, Fang Q, Minhe J. 1999. Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data. *International Journal of Remote Sensing* **20**:2805–2822 DOI 10.1080/014311699211804.

Johnson BA, Iizuka K. 2016. Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: case study of the Laguna de Bay area of the Philippines. *Applied Geography* **67**:140–149 DOI 10.1016/j.apgeog.2015.12.006.

Kanniah K, Sheikhi A, Cracknell A, Goh H, Tan K, Ho C, Rasli FN. 2015. Satellite images for monitoring mangrove cover changes in a fast growing economic region in southern peninsular malaysia. *Remote Sensing* **7**:14360–14385 DOI 10.3390/rs71114360.

Karatzoglou A, Smola A, Hornik K, Zeileis A. 2004. kernlab-an S4 package for kernel methods in R. *Journal of Statistical Software* **11**:1–20.

Kelley L, Pitcher L, Bacon C. 2018. Using Google Earth engine to map complex shade-grown coffee landscapes in Northern Nicaragua. *Remote Sensing* **10**:952 DOI 10.3390/rs10060952.

Kennel P, Tramon M, Barbier N, Vincent G. 2013. Canopy height model characteristics derived from airbone laser scanning and its effectiveness in discriminating various tropical moist forest types. *International Journal of Remote Sensing* **34**:8917–8935 DOI 10.1080/01431161.2013.858846.

Koh LP, Wilcove DS. 2008. Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters* **1**:60–64 DOI 10.1111/j.1755-263X.2008.00011.x.

Kou W, Xiao X, Dong J, Gan S, Zhai D, Zhang G, Qin Y, Li L. 2015. Mapping deciduous rubber plantation areas and stand ages with PALSAR and Landsat Images. *Remote Sensing* **7**:1048–1073 DOI 10.3390/rs70101048.

Kuhn M. 2015. Caret: classification and regression training. R package version 6.0-71. Astrophysics Source Code Library.

Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, R Core Team, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C. 2016. caret: classification and regression training. R package version 6.0-71. *Available at https://CRAN.R-project.org/package=caret*.

Lee JSH, Wich S, Widayati A, Koh LP. 2016. Detecting industrial oil palm plantations on Landsat images with Google Earth Engine. *Remote Sensing Applications: Society and Environment* **4**:219–224 DOI 10.1016/j.rsase.2016.11.003.

Lennox GD, Gardner TA, Thomson JR, Ferreira J, Berenguer E, Lees AC, Mac Nally R, Aragao LE, Ferraz SF, Louzada J, Moura NG. 2018. Second rate or a second chance? Assessing biomass and biodiversity recovery in regenerating Amazonian forests. *Global Change Biology* **9** DOI 10.1111/gcb.14443.

Li X, Chen W, Cheng X, Wang L. 2016. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. *Remote Sensing* **8**:514 DOI 10.3390/rs8060514.

Li L, Dong J, Njeudeng Tenku S, Xiao X. 2015. Mapping oil palm plantations in cameroon Using PALSAR 50-m orthorectified mosaic images. *Remote Sensing* **7**:1206–1224 DOI 10.3390/rs70201206.

Li Z, Fox JM. 2012. Mapping rubber tree growth in mainland Southeast Asia using time-series MODIS 250 m NDVI and statistical data. *Applied Geography* **32**:420–432 DOI 10.1016/j.apgeog.2011.06.018.

Li G, Lu D, Moran E, Hetrick S. 2011. Land-cover classification in a moist tropical region of Brazil with Landsat TM imagery. *International Journal of Remote Sensing* **32**:8207–8230 DOI 10.1080/01431161.2010.532831.

Liaw A, Wiener M. 2002. Classification and regression by random Forest. *R News* **2**:18–22.

Liu SS, Wei XL, Li DQ, Lu DS. 2017a. Examining forest disturbance and recovery in the subtropical forest region of zhejiang province using landsat time-series data. *Remote Sensing* **9**:479 DOI 10.3390/rs9050479.

Liu X, Wu P, Shao X, Songer M, Cai Q, Zhu Y, He X. 2017b. Spatiotemporally monitoring forest landscape for giant panda habitat through a high learning-sensitive neural network in Guanyinshan Nature Reserve in the Qinling Mountains, China. *Environmental Earth Science* **76**:589 DOI 10.1007/s12665-017-6926-9.

Lu D, Li G, Moran E, Dutra L, Batistella M. 2011. A comparison of multisensor integration methods for land cover classification in the brazilian amazon. *GIsci Remote Sens* **48**:345–370 DOI 10.2747/1548-1603.48.3.345.

Lu D, Li G, Moran E, Kuang W. 2014. A comparative analysis of approaches for successional vegetation classification in the Brazilian Amazon. *GIsci Remote Sens* **51**:695–709 DOI 10.1080/15481603.2014.983338.

Marselis SM, Tang H, Armston JD, Calders K, Labrière N, Dubayah R. 2018. Distinguishing vegetation types with airborne waveform lidar data in a tropical forest-savanna mosaic: a case study in Lopé National Park, Gabon. *Remote Sensing of Environment* **216**:626–634 DOI 10.1016/j.rse.2018.07.023.

Maryantica N, Lin C. 2017. Exploring changes of land use and mangrove distribution in the economic area of Sidoarjo District, East Java using multitemporal Landsat images. *Information Processing in Agriculture* **4**:321–332 DOI 10.1016/j.inpa.2017.06.003.

Midekisa A, Holl F, Savory DJ, Andrade-Pacheco R, Gething PW, Bennett A, Sturrock HJ. 2017. Mapping land cover change over continental Africa using Landsat and Google Earth Engine cloud computing. *PLOS ONE* **12**:e0184926 DOI 10.1371/journal.pone.018s4926.

Mitchard ETA, Saatchi SS, Lewis SL, Feldpausch TR, Woodhouse IH, Sonké B, Rowland C, Meir P. 2011. Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest–savanna boundary region of central Africa using multi-temporal L-band radar backscatter. *Remote Sensing of Environment* **115**:2861–2873 DOI 10.1016/j.rse.2010.02.022.

Mitchard ETA, Saatchi SS, White LJT, Abernethy KA, Jeffery KJ, Lewis SL, Collins M, Lefsky MA, Leal ME, Woodhouse IH, Meir P. 2012. Mapping tropical forest biomass with radar and spaceborne LiDAR in Lopé National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences* **9**:179–191 DOI 10.5194/bg-9-179-2012.

Morel AC, Saatchi SS, Malhi Y, Berry NJ, Banin L, Burslem D, Nilus R, Ong RC. 2011. Estimating aboveground biomass in forest and oil palm plantation in Sabah, Malaysian Borneo using ALOS PALSAR data. *Forest Ecology and Management* **262**:1786–1798 DOI 10.1016/j.foreco.2011.07.008.

Mukul SA, Herbohn J, Firn J. 2016. Tropical secondary forests regenerating after shifting cultivation in the Philippines uplands are important carbon sinks. *Scientific Reports* **6**:22483 DOI 10.1038/srep22483.

Ningthoujam R, Tansey K, Balzter H, Morrison K, Johnson S, Gerard F, George C, Burbidge G, Doody S, Veck N, Llewellyn GM. 2016. Mapping forest cover and forest cover change with airborne S-band radar. *Remote Sensing* **8**:577 DOI 10.3390/rs8070577.

Pal M, Mather PM. 2005. Support vector machines for classification in remote sensing. *International Journal of Remote Sensing* **26**:1007–1011 DOI 10.1080/01431160512331314083.

Paneque-Gálvez J, Mas JF, Moré G, Cristóbal J, Orta-Martínez M, Luz AC, Gueze M, Macia MJ, Reyes-Garcia V. 2013. Enhanced land use/cover classification of heterogeneous tropical landscapes using support vector machines and textural homogeneity. *International Journal of Applied Earth Observation and Geoinformation* **23**:372–383 DOI 10.1016/j.jag.2012.10.007.

Phompila C, Lewis M, Clarke K, Ostendorf B. 2014. Monitoring expansion of plantations in Lao tropical forests using Landsat time series. In: *Proceedings of the SPIE conference on land surface remote sensing II. Volume 9260*. Bellingham: SPIE DOI 10.1117/12.2068283.

Pimple U, Simonetti D, Sitthi A, Pungkul S, Leadprathom K, Skupek H, Som-ard J, Gond V, Towprayoon S. 2018. Google earth engine based three decadal landsat imagery analysis for mapping of mangrove forests and its surroundings in the trat province of Thailand. *Journal of Computer and Communications* **6**:e81443 DOI 10.4236/jcc.2018.61025.

Quegan S, Zeng T, Hu C, Uryu Y, Dong X. 2014. Regional tropical deforestation detection using ALOS PALSAR 50 m mosaics in Riau province, Indonesia. *Electronics Letters* **50**:547–549 DOI 10.1049/el.2013.4254.

R Core Team. 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. *Available at https://www.R-project.org/*.

Raczko E, Zagajewski B. 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing* **50**:144–154 DOI 10.1080/22797254.2017.1299557.

Ramdani F, Moffiet T, Hino M. 2014. Local surface temperature change due to expansion of oil palm plantation in Indonesia. *Climatic Change* **123**:189–200 DOI 10.1007/s10584-013-1045-4.

Rembold K, Mangopo H, Tjitrosoedirdjo SS, Kreft H. 2017. Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biological Conservation* **213**:234–242 DOI 10.1016/j.biocon.2017.07.020.

Rogan J, Franklin J, Stow D, Miller J, Woodcock C, Roberts D. 2008. Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. *Remote Sensing of Environment* **112**:2272–2283 DOI 10.1016/j.rse.2007.10.004.

Rwanga SS, Ndambuki JM. 2017. Accuracy assessment of land use/land cover classification using remote sensing and GIS. *International Journal of Geosciences* **8**:611–622 DOI 10.4236/ijg.2017.84033.

Sesnie SE, Finegan B, Gessler PE, Thessler S, Ramos-Bendana Z, Smith AMS. 2010. The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees. *International Journal of Remote Sensing* **31**:2885–2909 DOI 10.1080/01431160903140803.

Sesnie SE, Gessler PE, Finegan B, Thessler S. 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment* **112**:2145–2159 DOI 10.1016/j.rse.2007.08.025.

Shen W, Li M, Wei A. 2017. Spatio-temporal variations in plantation forests' disturbance and recovery of northern Guangdong Province using yearly Landsat time series observations (1986–2015). *Chinese Geographical Science* **27**:600–613 DOI 10.1007/s11769-017-0880-z.

Shiraishi T, Motohka T, Thapa RB, Watanabe M, Shimada M. 2014. Comparative assessment of supervised classifiers for land use–land cover classification in a tropical region using time-series PALSAR mosaic data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **7**:186–1199 DOI 10.1109/JSTARS.2014.2313572.

Silva J, Bacao F, Caetano M. 2017. Specific land cover class mapping by semi-supervised weighted support vector machines. *Remote Sensing* **9(2)**:181 DOI 10.3390/rs9020181.

Silva J, Bacao F, Dieng M, Foody GM, Caetano M. 2017. Improving specific class mapping from remotely sensed data by cost-sensitive learning. *International Journal of Remote Sensing* **38**:3294–3316 DOI 10.1080/01431161.2017.1292073.

Singh M, Evans D, Chevance JB, Tan BS, Wiggins N, Kong A, Sakhoeun S. 2018. Evaluating the ability of community-protected forests in Cambodia to prevent deforestation and degradation using temporal remote sensing data. *Ecology and Evolution* **8**(**20**):10175–10191 DOI 10.1002/ece3.4492.

Singh M, Evans D, Coomes DA, Friess DA, Suy Tan B, Samean Nin C. 2016. Incorporating canopy cover for airborne-derived assessments of forest biomass in the tropical forests of Cambodia. *PLOS ONE* **11**:e0154307 DOI 10.1371/journal.pone.0154307.

Singh M, Evans D, Friess D, Tan B, Nin C. 2015. Mapping above-ground biomass in a tropical forest in cambodia using canopy textures derived from google earth. *Remote Sensing* **7**:5057–5076 DOI 10.3390/rs70505057.

Singh M, Friess DA, Vilela B, Alban JDTD, Monzon AKV, Veridiano RKA, Tumaneng RD. 2017b. Spatial relationships between above-ground biomass and bird species biodiversity in Palawan, Philippines. *PLoS ONE* **12**:e0186742 DOI 10.1371/journal.pone.0186742.

Singh M, Tokola T, Hou Z, Notarnicola C. 2017a. Remote sensing-based landscape indicators for the evaluation of threatened-bird habitats in a tropical forest. *Ecology and Evolution* **7**:4552–4567 DOI 10.1002/ece3.2970.

Singh SK, Srivastava PK, Gupta M, Thakur JK, Mukherjee S. 2014. Appraisal of land use/land cover of mangrove forest ecosystem using support vector machine. *Environmental Earth Sciences* **71**:2245–2255 DOI 10.1007/s12665-013-2628-0.

Sirro L, Häme T, Rauste Y, Kilpi J, Hämäläinen J, Gunia K, De Jong B, Paz Pellat F. 2018. Potential of different optical and SAR data in forest and land cover classification to support REDD+ MRV. *Remote Sensing* **10**:942 DOI 10.3390/rs10060942.

Szuster BW, Chen Q, Borger M. 2011. A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography* **31**:525–532 DOI 10.1016/j.apgeog.2010.11.007.

Thanh Noi P, Kappas M. 2018. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors* **18**:18 DOI 10.3390/s18010018.

Tiwari LK, Sinha SK, Saran S, Tolpekin VA, Raju PLN. 2016. Forest encroachment mapping in Baratang Island, India, using maximum likelihood and support vector machine classifiers. *Journal of Applied Remote Sensing* **10**:016016 DOI 10.1117/1.JRS.10.016016.

Torbick N, Chowdhury D, Salas W, Qi J. 2017. Monitoring rice agriculture across Myanmar using time series Sentinel-1 assisted by Landsat 8 and PALSAR 2. *Remote Sensing* **9**:119.

Tottrup C. 2004. Improving tropical forest mapping using multi-date Landsat TM data and pre-classification image smoothing. *International Journal of Remote Sensing* **25**:717–730 DOI 10.1080/0143116031000159892 6.

Trisasongko BH, Panuju DR, Paull DJ, Jia X, Griffin AL. 2017. Comparing six pixel-wise classifiers for tropical rural land cover mapping using four forms of fully polarimetric SAR data. *International Journal of Remote Sensing* **38**:3274–3293 DOI 10.1080/01431161.2017.1292072.

Tsai YH, Stow D, Chen HL, Lewison R, An L, Shi L. 2018. Mapping vegetation and land use types in fanjingshan national nature reserve using google earth engine. *Remote Sensing* **10**:927 DOI 10.3390/rs10060927.

Venables W, Ripley BD. 1997. *Modern applied statistics with S-PLUS*. Second edition. New York: Springer Science & Business Media.

Wang C, Gao Q, Wang X, Yu M. 2015. Decadal trend in agricultural abandonment and woodland expansion in an agro-pastoral transition band in Northern China. *PLOS ONE* **10**:e0142113 DOI 10.1371/journal.pone.0142113.

Werbos PJ. 1994. *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting (Vol. 1)*. Hoboken: John Wiley & Sons.

Wilcove DS, Giam X, Edwards DP, Fisher B, Koh LP. 2013. Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. *Trends in Ecology and Evolution* **28**:531–540 DOI 10.1016/j.tree.2013.04.005.

Win RN, Reiji S, Shinya T. 2009. Forest cover changes under selective logging in the kabaung reserved forest, bago mountains, myanmar. *Mountain Research and Development* **29**:328–338 DOI 10.1659/mrd.00009.

Woodcock P, Edwards DP, Fayle TM, Newton RJ, Khen CV, Bottrell SH, Hamer KC. 2011. The conservation value of South East Asia's highly degraded forests: evidence from leaf-litter ants. *Philosophical Transactions of the Royal Society B: Biolological Sciences* **366**:3256–3264 DOI 10.1098/rstb.2011.0031.

Xie Z, Chen Y, Lu D, Li G, Chen E. 2019. Classification of land cover, forest, and tree species classes with ZiYuan-3 multispectral and stereo data. *Remote Sensing* **11**(**2**):164 DOI 10.3390/rs11020164.

Zeng Z, Estes L, Ziegler AD, Chen A, Searchinger T, Hua F, Guan K, Jintrawet A, Wood E. 2018. Highland cropland expansion and forest loss in Southeast Asia in the twenty-first century. *Nature Geoscience* **11**:556–562 DOI 10.1038/s41561-018-0166-9.