**GENERAL INFORMATION**

1. Title of Dataset: Past deforestation (2000-2018) and future deforestation probability (2019-2053) for Wallacea

2. Contact Information

 Name: Matthew J. Struebig

 Institution: University of Kent

 Address: Canterbury

 Email:M.J.Struebig@kent.ac.uk

 Name: Maria Voigt

 Institution: University of Kent

 Address: Canterbury, UK

 Email: M.Voigt@kent.ac.uk

3. Data produced: September 2020

4. Geographic location of data: Wallacea, Indonesia

5. Information about funding sources that supported the collection of the data:

This study was funded under the Newton Fund’s Wallacea Programme via the UK Natural Environment Research Council (NERC, NE/S007067/1) and the Indonesian Ministry for Research, Technology & Higher Education (Ristekdikti, NKB-2892/UN2.RST/HKP.05.00/2020 and 1/E1/KP.PTNBH/2019).

6. Purpose of the study:

We assessed patterns and drivers of forest loss and fragmentation across Wallacea, and used dynamic deforestation models to project future deforestation to 2053. This provides a valuable baseline from which to monitor Wallacea’s new development course, as Indonesia undergoes profound policy changes that will provide both challenges and opportunities for environmental governance and conservation.

DATA & FILE OVERVIEW

1. File List

A. Filename: forest\_2014\_18\_180m.tif

Short description: Primary forest cover in Wallacea in the year 2014 (coded as 1), forest loss between 200 and 2013 (coded as 0). Non-forest coded as -9999. This layer was used to train the deforestation model.

B. Filename: deforestation\_2014\_18\_180m.tif

Short description: Primary forest cover loss in Wallacea between the years 2014 and 2018 (coded as 1). All other pixels coded as -9999. This layer was used to train the deforestation model.

C. Filenames: summed\_deforestation\_probability\_20xx\_20xx\_Wallacea.tif:

Short description: Summed probability of deforestation for the 5-year intervals from 2019 to 2053. The binary maps of projected deforestation (n = 100) were aggregated into the summed probability of deforestation (i.e., if a pixel was deemed to be deforested in 50 of 100 iterations, it was assigned a summed deforestation probability of 50%).

2. Relationship between files:

Files A and B were used alongside data in Table 1 in related publication Voigt et al. 2021 (https://kar.kent.ac.uk/89330/) in the deforestation model to produce files C.

**Table 1** Predictors used in deforestation modelling, including their description, source and year

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Description | Source | Year |
| Forest cover and loss | Forest cover and loss previous to the calibration period (2001-2013) and in the calibration period (2014-2018) | Giri et al. (2011), Hansen et al. (2013), Margono et al. (2014) | 2000,2001-2013,2014-2018 |
| Slope | Slope in 2000 derived from the digital elevation model (30 m) | Farr et al. (2007) | 2000 |
| Fire activity | The average number of active fires per year (MODIS and VIIRS) as a proxy for fire proneness and agricultural activity. | MODIS Collection 6 NRT (2018), VIIRS 375m NRT (2018) | 2000-2018,2012-2018 |
| Accessibility | Accessibility from settlements, considering roads, slope and landcover (Deere *et al* 2020, Weiss *et al* 2018) | Populated places (World Resources Institute (WRI)), Ministry of Environment and Forestry, Republic of Indonesia (2013), Roads (WRI), Slope (Farr *et al* 2007) | 1990-2011 |
| Human population pressure | Local population pressure (Σ = 1) | Accessibility, Rose et al. (2018) | 1990-2017 |
| Main commodity | Distance to an Indonesian village (*Desa*) (includes human settlements and surrounding land mapped by the Indonesian Bureau of Statistics) which derives income from staple food agricultural, plantation agriculture, non-agricultural or fisheries commodities | Indonesian Bureau of Statistics (2018) | 2018 |
| Transmigrant settlements | Distance to settlements with an ethnic majority from outside of Wallacea | Indonesian Bureau of Statistics (2011), Indonesian Ministry of Environment and Forestry (2015) | 2011 |
| Mining | Exploration and production mining concessions (absence of mining concessions as reference) | WRI | 2017 |
| Land-use | Non-forest areas (APL), production forests (HP, HPK), and limited production forests (HPT). Protected forests (CA, HSAW, KSPA, SM, TN, TAHURA, TNL, TWA, TWA/HW, TWAL, TB) as reference areas | Ministry of Forestry (2010) | 2010 |

3. Additional related data collected that was not included in the current data package:

See Table 1

4. Are there multiple versions of the dataset? No

**METHODOLOGICAL INFORMATION**

1. Description of methods used for collection/generation of data:

Description of forest definition, processing and modelling procedure in Voigt et al. 2021 (https://kar.kent.ac.uk/89330/ ) and accompanying SI:

## Processing of deforestation model predictors

All layers used were converted to the Asia South Albers Equal Area Conic projection and resampled to the same extent and origin at 180 x 180 m pixel size (bilinear for continuous predictors, and nearest-neighbour resampling for categorical). All spatial manipulations were performed in Python (Python Software Foundation 2019), and aggregated, analysed and visualized in Python, R (R Core Team 2020) and ArcGIS Pro (Esri 2020).

## Forest definition

We defined deforestation as the annual loss of forest including mangroves (Giri *et al* 2011, Margono *et al* 2014). Primary forest was defined as mature natural forest with an extent >5 ha, and a natural composition and structure that has not been cleared in recent history (Margono *et al* 2014). The forest definition includes mainly tall evergreen dipterocarps growing on drylands or swamps, including peat-swamps, with closed canopies (>90% cover) and high carbon stock (above-ground carbon: 150 - 310 Mg C/Ha). Young forest regrowth, agro-forests, mixed gardens, scrublands, tree plantations, agricultural land and non-vegetated areas were excluded (Margono *et al* 2014). This definition of forest cover, comprising both intact and degraded types of primary forest in the year 2000, corresponds well (90.2% agreement) to the forest definition used by the Ministry of Forestry in the year 2000 (Margono *et al* 2014, Ministry of Environment and Forestry Republic of Indonesia 2018). For the purpose of this study, the forest definition was expanded to include mangrove forests (Giri *et al* 2011), amounting to an additional 880 km2 of coastal forest.

Forest loss, defined as the removal or mortality of tree cover, was based on the Tree Loss dataset (v1.6) developed at University of Maryland with Landsat time-series imagery (Hansen *et al* 2013). We have sought to minimized the inclusion of permanent or temporary forest loss within industrial plantations and small-holder agriculture by excluding the loss of tree cover within plantations, agro-forests, mixed gardens, regrowth or scrubland. Forested pixels were defined as having >70% tree canopy cover at the Landsat pixel (30 m resolution) scale. We used yearly measures of forest loss and aggregated forest cover and loss at a 180 m resolution using nearest-neighbour resampling, to minimize inclusion of short-term and small scale degradation and to facilitate data processing and modelling (forest\_2014\_18\_180m.tif and deforestation\_2014\_18\_180m.tif).

## Modelling framework

The model of forest loss for each province and state was adapted from Rosa *et al* (2013) and is based on *Ptrloss,x,t*, the probability that trees in a pixel *x* are lost in a time interval *t*. The probability of loss is defined as a logistic function:

| $$Ptrloss\_{x,t}=\frac{1}{1+exp^{-k\_{x,t}}}$$ | (1) |
| --- | --- |

in which *kx,t* can range from minus to plus infinity and *Ptrloss,x,t* from 0 to 1. We then used linear models to describe *kx,t* as a function of the predictor variables that affect forest loss at location *x* and time *t*.

Using a forward stepwise regression, a set of models was fitted to the observed forest loss data (2014 – 2018). Each model differed in the combination of predictor variables that define *kx,t*. The total number of models was depended on the number of predictors for the respective sub-regions and varied from 56 to 79. The models were fitted using ‘Filzbach’, a freely available library (https://github.com/predictionmachines/Filzbach), which uses a Markov Chain Monte Carlo (MCMC) sampling method to return a posterior probability distribution for each parameter. From this distribution, given a specific parameter combination ϴ, the posterior mean and credible interval was extracted. To estimate the parameters, the log-likelihood, a measure of the goodness of fit between the observations and the model predictions, is defined for a particular combination of variables:

| $$L\left(X|s,ϴ\right)=\sum\_{}^{}log\left(Ζ\_{x,t}Ptrloss\_{x,t}+\left(1-Ζ\_{x,t}\right)\left(1-Ptrloss\_{x,t}\right)\right)$$ | (2) |
| --- | --- |

in which *Ζx,t* is the observed forest loss at location *x* and time *t,* and *s* one of the models considered.

A cross-validation technique was used to assess the predictive power gained by adding variables to the model. This technique allowed us to check how accurately the model predictions compared to a randomly selected subset of 50% of the data that was not used to train the model. This cross-validation is necessary to find models that only comprise predictors with evident predictive ability. After successively adding the variable that resulted in the highest likelihood model, the overall best model (i.e. the one with the maximum test likelihood) was selected from the whole set of models for each province.

## Simulations

The simulation was based on recalculating equation (1) for each time-step, while using a slightly different set of parameter values at each iteration, thereby incorporating parameter uncertainty. These values were drawn from a Gaussian distribution resulting from the MCMC fitting, using the estimated mean and standard deviation for each parameter. As a result we received an updated *Ptrloss,x,t* for each individual pixel in each individual time period. We subsequently evaluated whether or not the respective pixel was lost, by drawing a random number from a uniform distribution between 0 and 1. We then classified the pixel as lost if the number was less than the probability of deforestation *Ptrloss,x,t*. This procedure was repeated for all seven time-steps and run multiple times (n = 100 iterations) to assess the uncertainty in model predictions over time. The different iterations were aggregated into the summed probability of deforestation and represent the fraction of simulation runs in which the forest in a pixel *x* was lost (summed\_deforestation\_probability\_20xx\_20xx\_Wallacea.tif). All predictor variables were treated as static predictors by the model (only one time-step was considered), apart from forest loss in the neighbourhood of a pixel, which was dynamically updated by the model in each time-step.

REFERENCES

Deere N J, Guillera‐Arroita G, Platts P J, Mitchell S L, Baking E L, Bernard H, Haysom J K, Reynolds G, Seaman D J I, Davies Z G and Struebig M J 2020 Implications of zero-deforestation commitments: Forest quality and hunting pressure limit mammal persistence in fragmented tropical landscapes *Conservation Letters* **13** e12701

Esri 2020 *ArcGIS Pro 2.x* (Redlands, CA: Environmental Systems Research Institute) Online: www.esri.com/en-us/arcgis/products/arcgis-pro/

Farr T G, Rosen P A, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E and Roth L 2007 The shuttle radar topography mission *Reviews of geophysics* **45**

Giri C, Ochieng E, Tieszen L L, Zhu Z, Singh A, Loveland T, Masek J and Duke N 2011 Status and distribution of mangrove forests of the world using earth observation satellite data *Global Ecology and Biogeography* **20** 154–9

Hansen M C, Potapov P V, Moore R, Hancher M, Turubanova S, Tyukavina A, Thau D, Stehman S V, Goetz S J, Loveland T R, Kommareddy A, Egorov A, Chini L, Justice C O and Townshend J R G 2013 High-Resolution Global Maps of 21st-Century Forest Cover Change *Science* **342** 850–3

Indonesian Bureau of Statistics 2011 Indonesia - Pendataan Potensi Desa 2008 Online: https://mikrodata.bps.go.id/mikrodata/index.php/catalog/80

Indonesian Bureau of Statistics 2018 Indonesia - Pendataan Potensi Desa 2014 Online: https://mikrodata.bps.go.id/mikrodata/index.php/catalog/599

Margono B A, Potapov P V, Turubanova S, Stolle F and Hansen M C 2014 Primary forest cover loss in Indonesia over 2000-2012 *Nature Clim. Change* **4** 730–5

Ministry of Environment and Forestry 2015 Indonesian land cover map 1990-2013 Online: http://webgis.menlhk.go.id:8080/pl/pl.htm

Ministry of Environment and Forestry Republic of Indonesia 2018 *The State of Indonesia’s Forests 2018* (Ministry of Environment and Forestry Republic of Indonesia)

Ministry of Forestry G D of P 2010 Landuse maps (provincial planning maps/Forest Land Use by Consensus maps (TGHK) Online: https://gis-gfw.wri.org/arcgis/rest/services/commodities/MapServer/13

MODIS Collection 6 NRT 2018 MODIS Collection 6 NRT Hotspot / Active Fire Detections MCD14DL Online: https://earthdata.nasa.gov/firms

Python Software Foundation 2019 *Python Language Reference version 3.7* Online: http://www.python.org​

R Core Team 2020 *R: A language and environment for statistical computing.* (Vienna, Austria: R Foundation for Statistical Computing) Online: https://www.R-project.org/

Rose A N, McKee J J, Urban M L and Bright E A 2018 LandScan 2017

VIIRS 375m NRT 2018 NRT VIIRS 375 m Active Fire product VNP14IMGT Online: https://earthdata.nasa.gov/firms

Weiss D J, Nelson A, Gibson H S, Temperley W, Peedell S, Lieber A, Hancher M, Poyart E, Belchior S, Fullman N, Mappin B, Dalrymple U, Rozier J, Lucas T C D, Howes R E, Tusting L S, Kang S Y, Cameron E, Bisanzio D, Battle K E, Bhatt S and Gething P W 2018 A global map of travel time to cities to assess inequalities in accessibility in 2015 *Nature* **553** 333–6

**DATA-SPECIFIC INFORMATION**

All files have the same geographic characteristics:

Driver: GTiff/GeoTIFF

Size is 10102, 10593

Coordinate System is:

PROJCRS["Albers",

 BASEGEOGCRS["WGS 84",

 DATUM["World Geodetic System 1984",

 ELLIPSOID["WGS 84",6378137,298.257223563,

 LENGTHUNIT["metre",1]]],

 PRIMEM["Greenwich",0,

 ANGLEUNIT["degree",0.0174532925199433]],

 ID["EPSG",4326]],

 CONVERSION["Albers Equal Area",

 METHOD["Albers Equal Area",

 ID["EPSG",9822]],

 PARAMETER["Latitude of false origin",-15,

 ANGLEUNIT["degree",0.0174532925199433],

 ID["EPSG",8821]],

 PARAMETER["Longitude of false origin",125,

 ANGLEUNIT["degree",0.0174532925199433],

 ID["EPSG",8822]],

 PARAMETER["Latitude of 1st standard parallel",7,

 ANGLEUNIT["degree",0.0174532925199433],

 ID["EPSG",8823]],

 PARAMETER["Latitude of 2nd standard parallel",-32,

 ANGLEUNIT["degree",0.0174532925199433],

 ID["EPSG",8824]],

 PARAMETER["Easting at false origin",0,

 LENGTHUNIT["metre",1],

 ID["EPSG",8826]],

 PARAMETER["Northing at false origin",0,

 LENGTHUNIT["metre",1],

 ID["EPSG",8827]]],

 CS[Cartesian,2],

 AXIS["easting",east,

 ORDER[1],

 LENGTHUNIT["metre",1,

 ID["EPSG",9001]]],

 AXIS["northing",north,

 ORDER[2],

 LENGTHUNIT["metre",1,

 ID["EPSG",9001]]]]

Data axis to CRS axis mapping: 1,2

Origin = (-955125.384435000014491,2374195.376991000026464)

Pixel Size = (180.000000000000000,-180.000000000000000)

Metadata:

 AREA\_OR\_POINT=Area

Image Structure Metadata:

 COMPRESSION=DEFLATE

 INTERLEAVE=BAND

Corner Coordinates:

Upper Left ( -955125.384, 2374195.377) (116d18'50.28"E, 5d41'38.81"N)

Lower Left ( -955125.384, 467455.377) (115d44'31.08"E, 10d52'49.01"S)

Upper Right ( 863234.616, 2374195.377) (132d51' 2.98"E, 5d40'11.30"N)

Lower Right ( 863234.616, 467455.377) (133d22' 4.42"E, 10d54'17.65"S)

Center ( -45945.384, 1420825.377) (124d34' 7.32"E, 2d49'49.28"S)

Band 1 Block=10102x1 Type=Int32, ColorInterp=Gray

 NoData Value=-9999

This Voigt\_et\_al\_21\_data\_readme file was generated on 28.07.21 by Maria Voigt