Carbon Storage Potential in Degraded forests of Kalimantan, Indonesia

António Ferraz[1,2]*, Sassan Saatchi[1,2], Liang Xu[2], Stephen Hagen[3], Jerome Chave[4], Yifan Yu[1], Victoria Meyer[1], Mariano Garcia[5], Carlos Silva[1,6], Orbita Roswintiart[7], Ari Samboko[7], Plinio Sist[8], Sarah Walker[9], Timothy R.H. Pearson[9], Arief Wijaya[10], Franklin B. Sullivan[11], Ervan Rutishauser[12,] Dirk Hoekman[13], Sangram Ganguly[14]

[1] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA.*
[2] *Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA.*
[3] *Applied GeoSolutions, Newmarket, NH, USA*
[4] *Laboratoire Evolution et Diversité Biologique UMR 5174, CNRS Université Paul Sabatier, Toulouse, France*
[5] *Department of Geology, Geography and Environment, University of Alcalá, 28801 Alcalá de Henares, Madrid, Spain*
[6] *Department of Natural Resources and Society, College of Natural Resources, University of Idaho, (UI), Moscow, ID, USA*
[7] *Indonesia National Institute of Aeronautics and Space (LAPAN) Jl. Lapan No. 70, Pekayon Pasar*
[8] *Cirad, UPR Forests and Societies, Univ. Montpellier, Campus International de Baillarguet, Montpellier, France*
[9] *Ecosystem Services Unit, Winrock International, Arlington, USA*
[10] *Forest and Climate Manager, World Resources Institute, Washington DC, USA*
[11] *Earth Systems Research Center, Institute of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, USA*
[12] *CarboForExpert,1248 Hermance, Switzerland*
[13] *Water Systems and Global Change Group, Wageningen University, Wageningen, The Netherlands*
[14] *NASA Ames Research Center/BAERI, Moffett Field, California, CA, USA.*

* Corresponding author: António Ferraz,

1. Supplementary Forest Sampling

1.1 Field Sampling

The field inventory systems have been established by different partners and therefore vary in size, shape and design (Table SI1). In general, *drylands* forests have a higher spatial variability than peat swamp forest and are sampled using larger plots compared with wetlands. There is plots with squared, circular or rectangular shapes. The location of individual plots within each field inventory site can follow either a random or systematic design to ensure unbiased inference of the mean and variance

of the population. Note that the location of the field inventory sites has been defined by different partners and do not follow any probabilistic design. We dispose of 104 field plots (82 over drylands and 22 over wetlands) with sizes varying from 0.01 ha to 0.25 ha. Five out of the 7 forest inventory sites are located over *drylands* and the remaining two over wetlands populated by peat swamp pole forest.

We applied two different strategies to calculate plot-level aboveground biomass (AGB). Over the inventory plots where the bole diameter has been measured for all trees (d>10 cm), the AGB have been estimated at the tree-level and the resulting values were summed up to the plot level. Otherwise, the plot-level AGB is estimated using a sub-sampling of all trees within the plots and the total amount is extrapolated to the plot level using a scaling factor (column Scaling factor, Table SI1).

To calculate the tree-level AGB_t (Kg), we used Chave *et al.* (2014) equation:

$$AGB_t = \exp(-1.803 - 0.976E + 0.976\,ln(wd) + 2.673\ln(d) - 0.0299(\ln(d))^2) \qquad \text{SI1)}$$

where d (cm) is the diameter of the bole measured at the diameter at breast height or above the buttress, wd ($g\,cm^{-3}$) the wood density that depends on the tree species, E is a measure of environmental stress defined as a function of temperature seasonality, precipitation seasonality and climatic water deficit at any location of the globe. In this work, we set $E = -0.1$. In case either wd is unknown or species have not been identified, we apply $wd = 0.57\,g\,cm^{-3}$ that is the average value for Kalimantan calculated using our field inventory. The plot level AGB density (Mgha⁻¹) is calculated by normalizing the total AGB by the area of the plots (column AGB mean, Table SI1).

Table SI1. Field inventory plots used to calibrate the lidar-AGB model and independent plots for validation purposes (denoted by *). D and W stand for drylands and wetlands, respectively. S (squared), C (circular) and R (rectangular) correspond to the shape of individual plots. Rand and Sys stand for random and systematic, respectively, and correspond to the strategy that was used to define the spatial distribution of the plots within the forest inventory site. MCH corresponds to the top of Mean Canopy Height that was derived from the lidar dataset (SI1.2).

Study site	Plot name	Land cover	Size (ha)	Shape	Sampling	Scaling factor	AGB mean (Mgha⁻¹)	MCH lidar (m)
Malinau	Cnv09.1	D	0.25	S	Rand	no	476.68	32.05
Malinau	Cnv09.2	D	0.25	S	Rand	no	406.42	36.11
Malinau	Cnv09.3	D	0.25	S	Rand	no	213.67	28.54
Malinau	Cnv09.4	D	0.25	S	Rand	no	389.01	32.16
Malinau	Cnv12.1	D	0.25	S	Rand	no	299.57	35.46
Malinau	Cnv12.2	D	0.25	S	Rand	no	483.61	35.24
Malinau	Cnv12.3	D	0.25	S	Rand	no	391.92	36.76
Malinau	Cnv12.4	D	0.25	S	Rand	no	436.31	38.27
Malinau	P01.1	D	0.25	S	Rand	no	290.53	35.40
Malinau	P01.2	D	0.25	S	Rand	no	513.50	37.55
Malinau	P01.3	D	0.25	S	Rand	no	154.30	29.32
Malinau	P01.4	D	0.25	S	Rand	no	432.47	35.53
Malinau	P02.1	D	0.25	S	Rand	no	295.55	33.07
Malinau	P02.2	D	0.25	S	Rand	no	300.42	32.77
Malinau	P02.3	D	0.25	S	Rand	no	308.01	30.00
Malinau	P02.4	D	0.25	S	Rand	no	207.53	26.08
Malinau	P06.1	D	0.25	S	Rand	no	312.99	33.89
Malinau	P06.2	D	0.25	S	Rand	no	170.70	29.65
Malinau	P06.3	D	0.25	S	Rand	no	184.36	28.67

Malinau	P06.4	D	0.25	S	Rand	no	317.36	33.30
Malinau	P07.1	D	0.25	S	Rand	no	189.83	26.77
Malinau	P07.2	D	0.25	S	Rand	no	307.94	34.32
Malinau	P07.3	D	0.25	S	Rand	no	295.00	29.86
Malinau	P07.4	D	0.25	S	Rand	no	468.20	35.96
Malinau	P12.1	D	0.25	S	Rand	no	237.02	32.76
Malinau	P12.2	D	0.25	S	Rand	no	464.72	37.47
Malinau	P12.3	D	0.25	S	Rand	no	743.28	42.51
Malinau	P12.4	D	0.25	S	Rand	no	355.58	37.99
Sumalindo IV	P1	D	0.25	S	Rand	no	259.93	27.26
Sumalindo IV	P2	D	0.25	S	Rand	no	204.52	28.06
Sumalindo IV	P3	D	0.25	S	Rand	no	306.14	29.85
Sumalindo IV	P4	D	0.25	S	Rand	no	342.14	31.29
Sumalindo IV	P5	D	0.25	S	Rand	no	198.23	30.39
Sumalindo IV	P6	D	0.25	S	Rand	no	303.16	34.56
Sumalindo IV	P7	D	0.25	S	Rand	no	451.71	39.78
Sumalindo IV	P8	D	0.25	S	Rand	no	118.82	25.77
Sumalindo IV	P9	D	0.25	S	Rand	no	175.65	25.89
Sumalindo IV	P10	D	0.25	S	Rand	no	99.61	19.18
Sumalindo IV	P11	D	0.25	S	Rand	no	136.95	22.71
Sumalindo IV	P12	D	0.25	S	Rand	no	100.43	19.28
Sindo Lumber *	Plot-1	D	0.25	S	Rand	no	203.12	23.54
Sindo Lumber *	Plot-2	D	0.25	S	Rand	no	261.90	26.29
Sindo Lumber *	Plot-3	D	0.25	S	Rand	no	255.91	28.17
Sindo Lumber *	Plot-4	D	0.25	S	Rand	no	284.70	29.19
Sindo Lumber *	Plot-5	D	0.25	S	Rand	no	237.82	24.73
Sindo Lumber *	Plot-6	D	0.25	S	Rand	no	321.78	31.56
Sindo Lumber *	Plot-7	D	0.25	S	Rand	no	301.12	30.73
Sindo Lumber *	Plot-8	D	0.25	S	Rand	no	421.48	38.53
Sindo Lumber *	Plot-9	D	0.25	S	Rand	no	308.88	31.47
Sindo Lumber *	Plot-10	D	0.25	S	Rand	no	262.40	25.21
Sindo Lumber *	Plot-11	D	0.25	S	Rand	no	195.56	21.76
Sindo Lumber *	Plot-12	D	0.25	S	Rand	no	231.38	26.70
Timberdana *	BioP1	D	0.125	C	Rand	yes	525.55	53.67
Timberdana *	BioP2	D	0.125	C	Rand	yes	260.05	25.70
Timberdana *	BioP3	D	0.125	C	Rand	yes	249.74	31.62
Timberdana *	BioP4	D	0.125	C	Rand	yes	541.00	42.09
Timberdana *	BioP5	D	0.125	C	Rand	yes	415.99	38.82
Timberdana *	BioP6	D	0.125	C	Rand	yes	298.98	39.94
Timberdana *	BioP7	D	0.125	C	Rand	yes	313.12	30.88
Timberdana *	BioP8	D	0.125	C	Rand	yes	270.74	30.40
Timberdana *	BioP9	D	0.125	C	Rand	yes	283.94	27.14
Timberdana *	BioP10	D	0.125	C	Rand	yes	498.83	35.10
Timberdana *	LiDAR	D	0.125	C	Rand	yes	272.05	25.63
Rodamas *	BP1_0	D	0.125	C	Rand	yes	430.09	36.94
Rodamas *	BP2_0	D	0.125	C	Rand	yes	353.43	35.17
Rodamas *	BP3_0	D	0.125	C	Rand	yes	394.67	34.397
Rodamas *	BP4_0	D	0.125	C	Rand	yes	96.32	26.00
Rodamas *	BP5_0	D	0.125	C	Rand	yes	879.75	44.08
Rodamas *	BP6_0	D	0.125	C	Rand	yes	333.18	34.62
Rodamas *	BP7_0	D	0.125	C	Rand	yes	580.54	46.07
Rodamas *	BP8_0	D	0.125	C	Rand	yes	383.41	31.08
Rodamas *	BP9_0	D	0.125	S	Rand	no	299.04	36.19
Rodamas *	BP10_0	D	0.125	S	Rand	no	644.48	48.19
Rodamas *	BP15_0	D	0.125	S	Rand	no	615.30	41.98
Rodamas *	BP16_0	D	0.125	S	Rand	no	479.92	40.45
Rodamas *	LIDAR_1	D	0.125	S	Rand	no	184.52	18.31
Rodamas *	LIDAR_2	D	0.125	S	Rand	no	542.62	43.09
Rodamas *	LIDAR_3	D	0.125	S	Rand	no	763.42	37.98
Rodamas *	LIDAR_4	D	0.125	S	Rand	no	961.80	41.43
Rodamas *	LIDAR_5	D	0.125	S	Rand	no	748.48	55.13
Rodamas *	LIDAR_6	D	0.125	S	Rand	no	271.45	32.49
Rodamas *	LIDAR_7	D	0.125	S	Rand	no	532.86	44.13
USFS I *	DS2	W	1	R (250x40)	Rand	yes	354.16	20.55
USFS I *	DS3	W	1	R (250x40)	Rand	yes	562.21	31.82

USFS I *	DS1	W	1	R (250x40)	Rand	yes	285.61	14.06
USFS II *	SEB2	W	1	R (250x40)	Rand	yes	254.19	18.741
USFS II *	SEB3	W	1	R (250x40)	Rand	yes	293.56	19.74
USFS III *	KTP1	W	1	R (250x40)	Rand	yes	68.21	4.80
Mawas	p2	W	0.1	R (50x20)	Sys	yes	240.22	18.90
Mawas	p3	W	0.1	R (50x20)	Sys	yes	239.47	18.56
Mawas	p4	W	0.1	R (50x20)	Sys	yes	295.98	20.88
Mawas	p5	W	0.1	R (50x20)	Sys	yes	235.22	18.91
Mawas	p6	W	0.1	R (50x20)	Sys	yes	181.32	15.44
Mawas	p7	W	0.1	R (50x20)	Sys	yes	280.772	18.77
Mawas	p8	W	0.1	R (50x20)	Sys	yes	212.68	17.50
Mawas	p9	W	0.1	R (50x20)	Sys	yes	223.59	17.81
Mawas	p10	W	0.1	R (50x20)	Sys	yes	169.839	14.18
Mawas	p11	W	0.1	R (50x20)	Sys	yes	168.609	15.01
Mawas	p12	W	0.1	R (50x20)	Sys	yes	177.34	17.26
Mawas	p13	W	0.1	R (50x20)	Sys	yes	223.38	17.09
Mawas	p14	W	0.1	R (50x20)	Sys	yes	140.98	15.18
Mawas	p15	W	0.1	R (50x20)	Sys	yes	124.78	14.51
Mawas	p16	W	0.1	R (50x20)	Sys	yes	169.34	16.80
Mawas	p17	W	0.1	R (50x20)	Sys	yes	177.59	15.21

The AGB of drylands forests ranges between 96.3 Mgha^{-1} and 961.8 Mgha^{-1} with an average of 354.3 Mgha^{-1}. If we considered only the larger plots (0.25 ha) those values equal 99.6 Mgha^{-1}, 743.3 Mgha^{-1} and 300.6 Mgha^{-1}, respectively. The results are summarized at the forest site level in Table SI2.

Table SI2. Main characteristics of field inventory sites used to calibrate the AGB models and to assess the wall-to-wall AGB map accuracy (denoted by *). Their location is provided in Figure 1a and characteristics of individual plots are given in Table SI1. In addition to the AGB mean, we provide the $\pm$ standard deviation and the 5th and 95th percentiles of the AGB distribution.

Study Site	Land Cover	# Plots	Size (ha)	WD Mean (g/cm^3)	AGB Mean (Mg/ha)	5th (Mg/ha)	95th (Mg/ha)
Malinau	Drylands	28	0.25	0.59	344.48 $\pm$130.24	175.48	503.03
Sumalindo IV	Drylands	12	0.25	0.57	224.77 $\pm$109.90	100.06	391.44
Mawas	Wetlands	16	0.1	0.56	203.82 $\pm$47.98	136.93	284.57
Sindo Lumber*	Drylands	12	0.25	0.69	273.84 $\pm$61.06	199.72	366.64
Timberdana*	Drylands	11	0.125	0.57	357.27 $\pm$ 114.91	254.90	533.27
Rodamas*	Drylands	19	0.125	0.57	357.27 $\pm$ 231.72	175.70	887.96
USFS*	Wetlands	6	1	0.56	302.99 $\pm$ 159.81	114.71	510.20

1.2 Lidar Sampling

We collected a total of 57 lidar flights to measure the forest structure across drylands (66,439 ha) and wetlands (44,250 ha) forests (Meledy et al., 2017). The location of the samples was determined using a probabilistic approach called Reverse Randomized Quadrat Recursive Raster approach (Theobald *et al.,* 2007). There is 29 flights (red dots in Figure 1a) of approximately 1000 ha (0.5 km X 20 km) that were collected based on a systematic random sampling approach following the VCS tool VT0005 (Tittmann *et al.,* 2015). The remaining 29 lidar flights were located over existing field inventory plots to allow development of the lidar-AGB model or were collected during transit between sites to add extra samples of vegetation structure (orange dots in Figure 1a).

Airborne lidar and high resolution digital color imagery were collected by PT Surtech Prima, a Jakarta, Indonesia-based company, between October and December 2014. The lidar dataset, which was collected at pulse densities ranging between 4 and 10 points per m^2, was provided as rectified LAS files (ASPRS 2015) in 500 × 500 meter tiles. The data achieved vertical accuracy of 0.08 m (standard error at 1 sigma) on clear ground. All data were supplied in Indonesian National Datum (DGN95) which is effectively the same as the WGS84 GPS satellite datum (ITRF 2000 Reference Frame). Vertical datum of the data is geoidal height generated from the EGM08 spheroid / geoid separation model.

Raw unclassified point clouds were processed using open source software developed as part of this study. The software, lidar2dems (https://github.com/Applied-GeoSolutions/lidar2dems), is a collection of open-source command line utilities built upon the Point Data Abstraction Library (PDAL, http://www.pdal.io/) and its associated dependencies for point classification and points2grid (https://github.com/CRREL/points2grid) for gridding returns into digital elevation models (DEMs). Classification of ground returns was performed using the classify script within lidar2dems. Points were classified using PDAL ground, which is built upon the Progressive Morphological Filter (PMF, Zhang *et al.,* 2003). The PMF parameters, slope and cell size, were tuned based on *a priori* tests we conducted in different terrain and vegetation coverage types. For ground classification, we used four broad classification groups that describe terrain and vegetation within sub-areas of each site (a. non-forest with flat terrain: slope 1, cell size 3; b. forest with flat terrain: slope 1, cell size 2; c. non-forest with complex terrain: slope 5, cell size 2; d. forest with complex terrain: slope 10, cell size 2). All tiles falling within each sub-area were merged then, the merged LAS files were classified using the parameterization from the classification group defined by the sub-area.

Merged and classified LAS files were passed to the lidar2dems script, which is used to develop different gridded products, namely Digital Terrain Models (DTM) and Digital Surface Models (DSM) and Canopy Height Models (CHM) representing the height of the forest. Lidar returns for each LAS file were filtered to select those corresponding to ground in order to generate DTM at 1 m resolution. Each DTM was generated using all ground-classified points in an XML pipeline within PDAL and was dependent on points2grid. For gridding returns in points2grid, we used inverse distance weighting of points within each grid cell search radius. Grid cells with no ground returns were progressively filled using a gap-filling technique, which iteratively increases the search radius used in points2grid and fills no data pixels with the smallest available search radius for each pixel. Gaps were filled using search radii of 0.56 m (the circumscribing circle radius of a 1-m grid cell), 1.41 m (the radius of the circumscribing circle of a 2-m grid cell), 2.5 m, and 3 m, and remaining missing data were filled using nearest neighbor interpolation. DSM (1 meter resolution) were generated using all non-ground returns in points2grid with a search radius of 0.56 m, using the maximum return value within each grid cell search radius. The CHM (1 meter resolution) was calculated as the difference between the DSM and the DTM. Each DEM was clipped to sub-area extents using Gippy (https://github.com/gipit/gippy) and merged with alike DEM types within the site extent. The resultant DTM and DSM data were used to calculate Canopy Height

Models (CHM) that provide us reliable measurements on the canopy height vegetation structure variability.

Finally, we define the mean top of canopy height (MCH) that is a plot-level metric calculated by averaging the CHM (1 m resolution) within the boundaries of each plot. This metric represents the forest structure variability across the lidar scenes. The MCH for plot of the current study is show in the last column of Table SI1 and it is used to calibrate the lidar-AGB model (Eq. 1 in Section 2 and SI1.3).

1.3 Lidar-AGB Model and lidar-AGB sampling

The model to estimate AGB per unit area ($Mgha^{-1}$) relates field estimated AGB (Kg) to the estimated forest structure (given by lidar-derived metric MCH, m) according to a given plot area (ha). A power law better fits the AGB-MCH relationship (Eq. 1), which agrees with models developed over other regions (Saatchi *et al.* 2011, Asner and Mascaro 2014, Xu *et al.* 2017). The lidar-AGB model has been linearized to estimate the parameters *a* and *b* of Eq. 1 and their residuals were analyzed graphically and tested for normality as well as for heteroscedasticity. The errors and uncertainty of the lidar-AGB model are assessed using a bootstrapping approach in SI 6.2.

The field plots available limit us to establish the lidar-AGB models at 0.25 ha and 0.1ha for *drylands* and wetlands, respectively (Figure 3). We selected the resolution for which we had the largest amount of field plots available in order to guarantee the best characterization of the landscape in terms of AGB variability. As far as *drylands* is concerned, we used the sites of Malinau and Sumalindo IV whereas the wetlands model is calibrated using Mawas. The data used to calibrate Eq. 1 are shown in the two last columns of Table SI1. Although the Sindo Lumber inventory is located in drylands and it is composed of plots with 0.25 ha, this site has not been used to develop the lidar-AGB model because it would underestimate the AGB findings (mainly of the high-density areas) and we consider that Malinau and Sumalindo IV better represent the overall Kalimantan landscape. Then, the calibrated lidar-AGB models are applied to the lidar raster product in order to convert lidar measurements on forest structure into AGB sampling datasets. Due to the fact that the lidar-AGB models have been established using different spatial resolutions, we compute two raster products with different cell sizes by resampling the original 1 m CHM. The lidar-derived CHM is resampled to 0.25 ha (50 m x 50 m) and 0.1 ha (~31.6 m x ~31.6 m) for *drylands* and wetlands, respectively. To do this, we define two empty grids with origin (center of the top left pixel) in the WGS/84 coordinates 8° *N* 108° *E* and with an extent that covers the area of interest (refer to SI2.3 for details on the grid). Then, the empty grid cells are filled up using the average of the 1 m CHM pixels located within it. Grid cells that have less than 90% of its area covered by the 1m pixels are left empty to avoid the edge effect. After applying the lidar-AGB models (Figure 3) to the corresponding MCH products, the latter are resampled to a grid with 1 ha pixel size by averaging the AGB pixels within each grid cell. Again, to avoid edge effect we do not consider the 1 ha cells which area is not covered by more than 90%

of the finer resolution pixels. Finally, the two 1 ha grids are combined into a single product by selecting the pixels cover. Note that the raster products overlap at the pixel-level because they have been resampled according to a common grid (SI2.3). The final product provides 110,689 points of AGB sampling density within areas of 1 ha.

1.4 Lidar Forest Degradation Index (FDI)

Kalimantan has been impacted by several wood extraction activities that include illegal logging, selective logging and harvesting. Here, we develop a wall-to-wall map of a forest degradation index (1 ha resolution) that allows to derive spatial explicit information on the intensity of logging activities and their impact on AGB density across Kalimantan landscape. To do this, we define the Forest Degradation Index (FDI) as shown in Eq. 2, i.e. $FDI = MCH + LCR + PC$ where MCH, LCP and PC stands for top of mean canopy height, large crown trees percentage, and percentage cover, respectively. Therefore, the FDI map results from three different wall-to-wall maps (1 ha resolution) that have been calculated separately using the Random Forest (RF) algorithm together with the remote sensing satellite layers used to calculate the AGB map (refer to SI3). In the following, we explain the method used to calculate the lidar-derived sampling training data required by the RF algorithm to calculate MCH, LCR and PC.

As far as the MCH is regarded, the cells of the 1 ha grid are calculated by averaging the original 1 m CHM. The PC is computed by calculating the percentage of pixels of the original CHM that are higher than 5 meters within each 1 ha cell. This height threshold agrees with the guidelines of the CDM of the Kyoto Protocol (UNFCCC, 2002) regarding the definition of forest, i.e. an area covered by vegetation higher than 5 meters and with PC larger than 30%. Finally, the LCR is the percentage of area covered by large trees within each 1 ha grid cell. This area is estimated by applying the LCA (Large Crown Area) algorithm to the 1 m CHM (Meyer *et al.*, 2018). Briefly, the LCA sets a threshold of 27 m on the 1 m CHM to derive a binary image representing the areas covered by high vegetation. Then, it applies a connected component segmentation (with an 8-neighborhood pixel connectivity) to compute clusters of pixels by gathering together all the neighbor's pixels. Finally, the clusters composed of less than 100 pixels (i.e. 100m^2) are removed. Therefore, the LCA technique provides a binary image at 1 m resolution that is resampled to a 1 ha grid representing the percentage of area covered by large trees. Note that the three products have been calculated considering the rule of the 90% covered pixels described in SI1.3 to avoid the edge effect.

2. Supplementary Satellite Remote Sensing Products

2.1. Raster

In this work, we use four medium resolution raster products as predictors within the Random Forest (RF) algorithm. The first one is a mosaic of Landsat 8 top-of-atmosphere (TOA) reflectance data (https://landsat.gsfc.nasa.gov/landsat-8/) averaged for the years of 2013 to 2016. We used the Google Earth Engine (https://earthengine.google.com/) to remove clouds and calculate the average of valid pixels for those 4 years of observation. The final mosaic includes the most meaningful bands for vegetation mapping purposes that are used by the well known Global Forest Change map provided by the University of Maryland (Hansen *et al* 2013). It includes band 3 (Red), 4 (NIR), 5 (SWIR-1) and 7(SWIR-2) at 30 m spatial resolution. The second raster product is the Digital Elevation Model (DEM) derived from the NASA Shuttle Radar Topography Mission (SRTM, http://www2.jpl.nasa.gov/srtm/) that characterizes the surface using a combination of C-band and X-band radar measurements. We used the version SRTM NASA V3 version defined at 30 m of resolution published in 2015. Data gaps due to radar layover and shadow, which are found especially in rugged terrain in mountainous areas (Sephton and Wishart 2013), are interpolated using ASTER GDEM2 (Global Digital Elevation Model Version 2). The third product is a radar backscattered raster derived from measurements of the Phased Array L-band Synthetic Aperture Radar (PALSAR, http://global.jaxa.jp/projects/sat/alos2) sensor carried on board of the Advanced Land Observing Satellite (ALOS), which is managed by the Japan Aerospace Exploration Agency (JAXA). We use the ortho- and slope-corrected backscatter coefficient in both HH and HV polarizations defined at 25 m resolution. The fourth input is also satellite radar backscatter product managed by the European Space Agency (ESA) that uses C-band SAR observations in the framework of the mission Sentinel-1 (https://sentinel.esa.int/web/sentinel/missions/sentinel-1). We used both polarizations (HH and HV) at 25m spatial resolution. The radar products are complementary because L- and C-band are sensitive to different elements of forest structure such as leave plant material or woody volume.

Finally, we use an additional raster product outside the framework of RF mapping to analyze results over wetlands in terms of AGB density as explained in SI4. This land cover map provided by (Hoekman *et al.,* 2010) classifies the Borneo territory according to the Land Cover Classification System (LCCC) of the Food and Agriculture Organization of the United Nations (FAO, Di Gregorio and Jansen 2000). The classification applies a Bayesian approach based on a mixture modeling followed by a Markov Random Field (MRF) to a ALOS-PALSAR backscatter raster product. We are interested in the delineation of peat swamp pole forest, peat swamp pandang forest, alang-alang, burn peat swamp forest and riverine forest (SI4).

2.2 Vector

We use three vector maps in this work. The first one is a land cover map derived from the Indonesian Ministry of Forestry (IMF) land cover map that was calculated using visual interpretation of Landsat images (SNI 2016). The multiple activities around the natural resources of Kalimantan (e.g. forest, oil, gas, coal and other minerals) and large-scale land use for crops (e.g. rice) and agroforestry plantations (e.g. oil palm and wood fibre) have converted Kalimantan into a fragmented landscape composed of a mosaic of natural vegetation and different land use units. The original 26 land cover classes have been aggregated into 9 classes as show in Figure SI1. The classes crops/agriculture and urban/settlement are not considered in this work for AGB estimation purposes. The characteristics of the crops/agriculture class are seasonal whereas the remote sensing products are not sensitive to the urban/settlement vegetation namely due to the coarse spatial resolution. Intact montane forest refers to intact forest above 1000 meters asl (above sea level). The IMF map plays three different roles within this work. First, it is used as a predictor layer within the RF algorithm (SI3). To do this, we apply a tool called gdal_rasterize (http://www.gdal.org/gdal_rasterize.html) from the GDAL to convert the vector file into a raster product with a resolution of 25 m to compare with most of the satellite raster products and then the raster is resampled according to the 1 ha common grid (SI2.3). The second role is to compute the drylands/wetlands mask show in Figure 1a, which allows to define the lidar-AGB models specific oriented to the vegetation type. Wetlands are defined by aggregating the classes peat swamp forest and swamp scrublands; whereas the remaining ones are considered as *drylands* (Figure SI1). Finally, the IMF map is as well used to analyze the results. It allows to aggregate the wall-to-wall AGB map results by land cover class and analyze the AGB distribution according to different forest types (Figure 4b).

The second vector map represents the administrative boundaries of the five Kalimantan regions (Figure SI5, http://www.gadm.org/), which is overlapped to the wall-to-wall AGB map to report AGB storage at the region level. With a total area of 54 Mha, it is divided into five provinces, namely North, East, South, Central and West Kalimantan. Finally, the third vector map delineates the selective logging concessions, the palm oil concessions and the wood fibre concessions. They are available in the Global Forest Watch website from Greenpeace (GFW, 2017) and are used here to report the AGB density by concession type (Figure SI4).
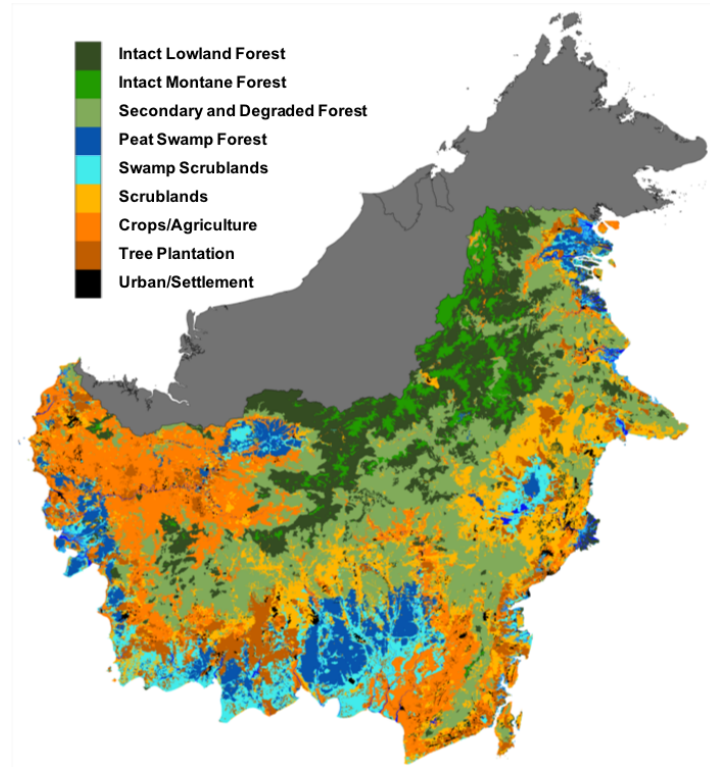
Figure SI1. Map derived from the Indonesian Ministry of Forestry (IMF) land Cover product (SNI 2016). The classes crops/agriculture and urban/settlement are not considered for AGB estimation purposes.

2.3 Resampling of the data to a common grid

The remote sensing products have been acquired by different sensors at different resolutions and therefore they cannot be directly compared. We decided to resample all satellite and airborne datasets to a spatial grid with 1 ha (0.00090109°) of spatial resolution. The grid is defined in the WGS 84 coordinate reference system with origin (center of the top left pixel) over 8° *N* 108° *E* and it has 13317 columns and 14427 rows to cover the entire Borneo. Obviously, many of these pixels will not be used because they cover either ocean or the non-Indonesian territory of Borneo. The resampling of the datasets was performed using the *gdalwarp* (http://www.gdal.org/gdalwarp.html) tool of the library Geospatial Data Abstraction Library (GDAL) using the "average" metric to upscale the datasets, with exception of the categorical IMF map that has been resampling using the metric "mode" to process existing the land cover classes.

3. Supplementary Random Forest Geospatial Modeling

Non-parametric machine learning regression algorithms have been widely used in the framework of different remote sensing applications over urban and natural landscapes (Guo *et al.,* 2011, Ferraz *et al.,* 2016). Its ability to map the AGB spatial distribution has been show at the regional-, national or global-scale (Saatchi *et al.,* 2011, Mascaro *et al.,* 2014, Baccini *et al.,* 2012). We apply the Random Forest (RF) machine learning algorithm after comparing its performance with the Maximum Entropy method using a set of training and test datasets. RF is an extension of the decision trees approach that has been developed to overcome the issue of overfitting to the training data when trees grown very deep that lead to highly irregular patterns with low bias but very high variance (Breiman 2001). Fitting multiple deep decision trees using randomly selected different parts of the training dataset and calculating its average highly reduces the variance. This comes at the expense of a small increase in the bias and some loss of interpretability. Whereas the bias can be easily corrected (Xu *et al.,* 2017), the loss of interpretability is a major drawback of the RF. Briefly, RF tends to predict toward the mean of the original training dataset giving accurate predictions on the mean of a given distribution but it faces difficulties in characterizing its tails. The impact of this issue in our AGB estimates is addressed in (Section 4.2 and SI6.3). We use the Matlab function *TreeBagger* that allows combining the results of multiple decision trees in order to minimize the overfitting (Breiman 2001). The parameters of the *TreeBagger* function, which were calibrated and evaluated using training and test datasets, include the total number of trees to grow, the minimum number of observations per leaf and, the number of variables to select at random for each decision split. They were set to 500, 5 and, 4, respectively.

Here, we use a single dataset of independent variables (or predictors) that are the remote sensing products described in SI2, namely the Landsat LC8, ALOS PALSAR, Sentinel-1, SRTM data and the IMF land cover map. In the following, we describe the lidar-derived training datasets (i.e. depend variables) required by the RF to calculate the wall-to-wall maps.

3.1 Mapping of Aboveground Biomass and Forest Degradation Index

We calculate four wall-to-wall maps using four different lidar–derived sampling datasets. The maps are called AGB (aboveground biomass), the MCH (top of Mean Canopy Height), the LCR (Large Crown Ratio) and the PC (Percentage Cover). The training data for the AGB map is described in SI1.3, whereas details for the remaining ones can be found in SI1.4. The AGB map is the main outcome of this work and it is show in Figure 4a. The MCH, LCR and PC maps are used to calculate the Forest Degradation Index (FDI) map (Figure SI2) by means of Eq. 2, i.e. FDI=MCH+LCR+PC. Both LCR and PC are defined in percentage whereas the MCH have been normalized to range within the interval [0-100] in order to compare with the remaining products.

The FDI map is examined together with high resolution images (namely, the 1 m CHM and the original satellite images described in SI2) to define the following degradation classes: intact forest (FDI>255), light degraded forest (235<FDI<=255), moderate degraded forest (210<FDI<=235), high degraded forest (150<DI<=210) and severe degraded forest (FDI<=150). The spatial distribution of the aforementioned classes is show in Figure 4c.
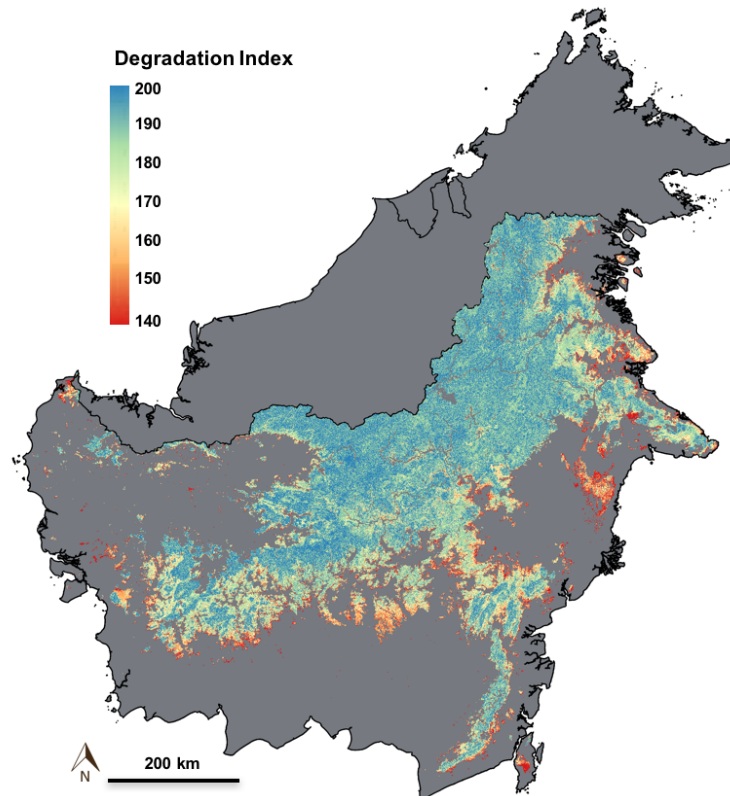
Figure SI2. Forest Degradation Index (FDI) map.

4. Supplementary carbon in wetlands

Kalimantan has nearly 7.58 Mha of wetlands that play a major role in maintaining the biodiversity and the regional hydro-ecological functions. The extent and structure of the forest wetlands, particularly across the peatlands, have gone through major changes due to deforestation and drainage for land use activities (Miettinen *et al.,* 2016). We examined the existing conditions of wetlands by analyzing the distribution of AGB across different forest types. The uncertainty of AGB estimates over all wetlands depends strongly on the uncertainty associated with the IMF land cover classification for delineating all peatland and coastal wetland vegetation types. In particular, it will have a significant impact on the AGB estimates uncertainty of

vegetation types with limited spatial extent (e.g. primary peat swamp pole forest). By using high resolution satellite data (30 m Landsat) and the 1-m lidar data acquired over the wetland areas, we were able to visually examine and edit inconsistencies of the land cover map and minimize the uncertainty in estimating AGB in these forests. We decided to report the AGB results directly from the lidar-derived maps instead of the RF predictions due to two main reasons. On the one hand, it would be unfeasible to edit the land cover map in detail over the wetlands covered by the wall-to-wall AGB map. On the other hand, we remove from the analysis the uncertainty associated with the RF mapping.

The wetlands forest types of interest are the primary peat swamp tall pole forests, secondary peat swamp tall pole forests, peat swamp pandang forests, burnt peat swamp forest, riverine forests, alang-alang. To complement the IMF map (SNI 2016), which does not detail many of these types of forest, we use the land cover map provided by (Hoekman *et al.,* 2010) described in SI2.1. Briefly, we selected riverine forest, peat swamp pandang forest, alang-alang and burnt peat swamp forest from Hoekman's and colleagues map. The remaining classes (primary peat swamp tall pole forest, secondary peat swamp tall forest) are selected from the IMF product. For coastal wetlands, including mangroves and nipah palm forests, the airborne lidar samples were limited in extent and could not provide reasonable estimates.

5 Supplementary results

5.1. AGB estimations of highly dense forests

Remote sensing signals in the optical and SAR wavelengths are known to flatten off and saturate over forest with high AGB density. However, our RF approach produces unbiased AGB estimates even for dense forests (> 400 Mgha^{-1}, Figure 4, 5 and SI7a). In our analysis, high AGB values correspond to areas with higher elevation following the pattern of the SRTM, which is the most meaningful variable for our mapping (Figure SI3). If this was not the case, we believe that RF algorithm would underestimate the AGB of dense forests. We believe that the large uncertainty on the pixels corresponding to high AGB forests is largely due to random errors (e.g. the prediction largely depends on a single layers) and less due to systematic errors (e.g. saturation).

There are also other layers that have contributing to extrapolating high AGB forests from the lidar on the RF map namely on lowlands where most of forests are degraded. These include optical imagery such as Landsat short wavelengths (SWIR1 & 2) that help to separate layered and shadowed forests in heterogeneous old growth forests from homogeneous and less layered secondary and degraded forests.
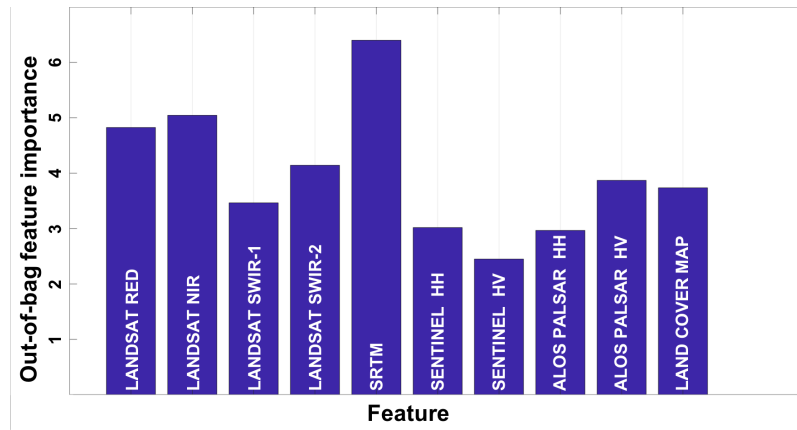
Figure SI3. Feature importance on the AGB calculation using the RF algorithm.

5.2. AGB density per concession areas

The results of AGB density per concession type show in Table SI3 have been calculated by overlapping the shapefiles described in SI2.2 to the AGB density map (Figure 4a). The overspread across the territory of the concession areas is show in Figure SI4.

Table SI3. Kalimantan's averaged AGB density $\pm$ the corresponding uncertainty calculated considering the spatial autocorrelation (SI6.4, Equation SI2). The 5th and 95th columns correspond to the percentiles of the AGB density distribution.

Land Cover	Surface covered (Mha)	AGB density (Mgha^{-1})	5th	95th
Selective Logging	10.6	258.0 ± 16.5	40.61	440.8
Palm Oil	9.82	59.7 ± 11.3	0.9	276.1
Wood Fibre	5.5	95.4 ± 13.2	2.4	318.4

5.3. AGB density by administrative region

The AGB density and storage within each administrative region of Kalimantan is valuable information to define forest and AGB management policies or emission reduction projects at the region level (Figure SI5a, Galudra *et al.,* 2011). Also, we provide results regarding intact forest, secondary and degraded forest, peat swamp forest and others for each region (Figure SI5b and c, Table SI4). Intact forest has been calculated by aggregating intact lowland and montane forest, whereas the class others comprise swamp scrublands, scrublands and tree plantations (Figure SI1).
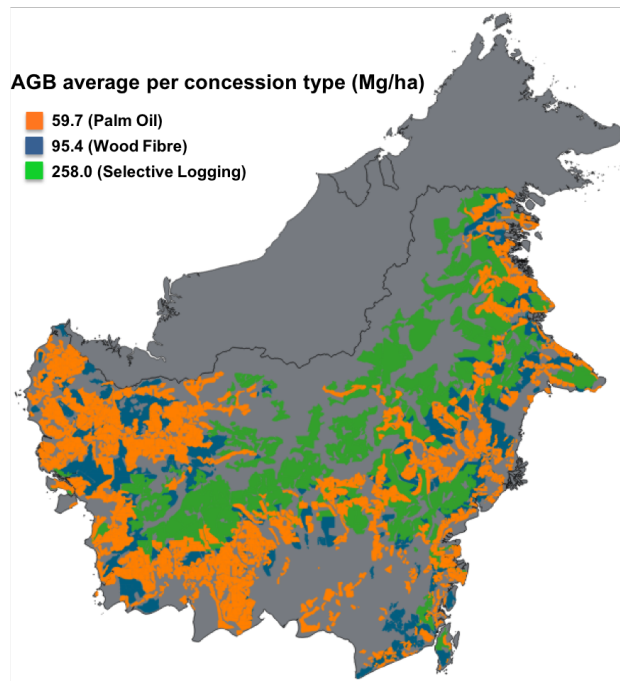
Figure SI4. Map of averaged AGB per concession area.

North Kalimantan has the highest AGB density because it has the highest ratio of intact forest to total area (280.5 $\pm$ 16.3 Mgha^{-1}, Table SI4 and Figure SI5) most likely due to the rough topography of that region. It is followed by East Kalimantan in terms of density (218.4 $\pm$ 7.6 Mgha^{-1}), which stores the largest amount of AGB of Kalimantan mainly due to the fact that it is the biggest region in terms of surface covered by vegetation and, it preserves a large area covered by *drylands* forest in a good balance between intact and degraded Forest (3.64 Pg, Figure SI5c). Both West and Central Kalimantan are also large in surface but they store less AGB because its landscape, which is characterized by a gentle topography, has been converted into logging or agricultural areas (2.27 Pg and 2.92 Pg, respectively). Likely North Kalimantan, West Kalimantan stores more AGB on intact forest than on secondary and degraded forest. South Kalimantan only accounts for 3% (0.38 Pg) of the Kalimantan's AGB and has almost no intact AGB mainly due to its small surface located in lowlands near the coastal areas.

The average and distribution of AGB density for each land cover class do not change significantly across regions (Figure SI5b). There are some exceptions (e.g. intact forest and peat swamp forest in South Kalimantan or peat swamp in East Kalimantan) because they cover small regions and might not be statistically representative. Peat swamp forests areas are mainly represented in Central and West Kalimantan. Besides its natural conditions, these regions contain some conservation areas (e.g. lake Sentarum and north part of Sebangau National Park) that helped to preserve this endemic land cover class. The remaining regions store almost no peat swamp AGB.

Table SI4. AGB average density by region and main cover classes. The $\pm$ error regards the uncertainty with respect to average estimator considering the spatial autocorrelation (SI6.4 and Equation SI2).

Regions	Surface covered (Mha)	AGB density Mean (Mg ha⁻¹)	5ᵗʰ	95ᵗʰ	AGB stock (Pg)
Central Kalimantan	**16.02**	**182.5 $\pm$ 2.2**	**1.9**	**414.5**	**2.92**
Intact Forest	2.14	312.6 $\pm$ 21.4	136.8	448.168	0.57
Secondary and Degraded Forest	6.14	261.3 $\pm$ 24.4	80.0	448.2	1.41
Peat Swamp Forest	2.05	214.5 $\pm$ 20.3	42.9	383.1	0.47
Others	6.32	38.0 $\pm$ 6.5	0.5	139.6	0.27
East Kalimantan	**16.67**	**218.4 $\pm$ 7.6**	**5.2**	**433.9**	**3.64**
Intact Forest	5.20	313.7 $\pm$ 18.4	137.6	449.5	1.40
Secondary and Degraded Forest	6.60	267.5 $\pm$ 24.0	77.7	458.2	1.53
Peat Swamp Forest	0.14	121.7 $\pm$ 12.1	10.4	299.6	0.02
Others	5.10	48.8 $\pm$ 7.2	1.1	182.5	0.29
North Kalimantan	**8.81**	**280.5 $\pm$ 16.3**	**31.8**	**443.5**	**2.47**
Intact Forest	5.12	319.0 $\pm$33.5	197.3	443.7	1.40
Secondary and Degraded Forest	2.79	268.4 $\pm$ 22.6	84.3	461.0	0.65
Peat Swamp Forest	0.25	188.9 $\pm$ 7.6	31.0	306.5	0.05
Others	0.71	68.2 $\pm$ 7.7	1.6	269.4	0.06
South Kamilantan	**2.47**	**143.2 $\pm$ 5.8**	**1.1**	**408.1**	**0.35**
Intact Forest	0.12	185.8 $\pm$ 19.8	8.9	396.9	0.02
Secondary and Degraded Forest	1.06	264.7 $\pm$ 23.1	67.8	460.6	0.25
Peat Swamp Forest	0.02	135.6 $\pm$ 9.5	8.0	245.0	0.00
Others	1.46	39.0 $\pm$ 7.5	0.5	160.2	0.06
West Kamilantan	**10.14**	**224.7 $\pm$ 4.0**	**2.0**	**445.6**	**2.27**
Intact Forest	3.15	332.5 $\pm$ 27.8	200.6	460.1	0.88
Secondary and Degraded Forest	3.32	264.5 $\pm$ 25.0	83.6	465.46	0.77
Peat Swamp Forest	1.20	223.0 $\pm$ 18.1	24.9	455.3	0.28
Others	3.10	35.6 $\pm$ 7.3)	0.3	148.7	0.11

6. Supplementary Uncertainty assessment

The uncertainty of the lidar-AGB model and the RF predictions were evaluated using a cross-validation bootstrapping approach by randomly selecting 70% of samples for modelling and 30% for validation (SI6.1 and SI6.2). The RF bootstrapping approach provides a series of AGB maps that allow for the calculation of a pixel-level uncertainty map capturing the stability of our model predictions spatially across gradients of forest structure and surface topography (SI6.3). The errors identified in SI6.1 and SI6.3 are propagated throughout the wall-to-wall map to assess the uncertainty in estimating the average AGB for large regions such as land cover types and administrative regions SI6.4. Finally, we compare the RF map directly with field AGB estimates (SI6.5).
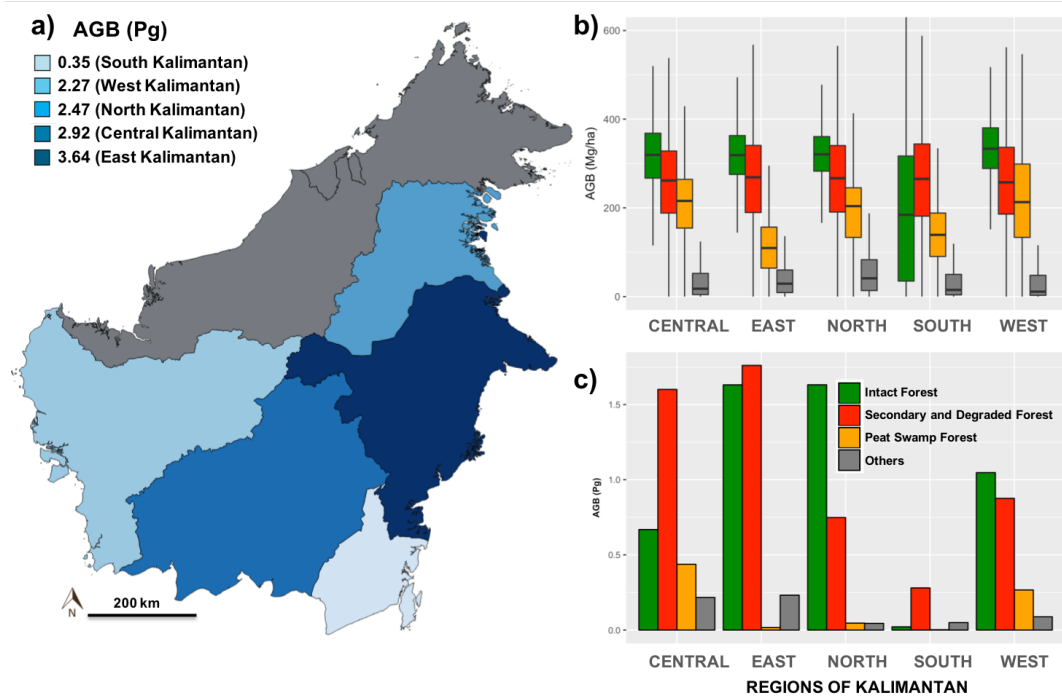
. Storage of AGB **a)** by region and **b)** its distribution and **c)** storage by land cover class.

6.1 Lidar-AGB model

We evaluate the relationship and errors associated with the estimates of the lidar-AGB models (both *drylands* and wetlands, Figure 3) using a bootstrapping (1000 times) cross-validation approach by applying the *boot* function from the programming language R package "boot" (Canty and Ripley 2017). The error of the drylands forest model is of RMSE=62.21 Mgha^{-1} and it is relativity unbiased (-2.03 Mgha^{-1}) with a standard error (standard deviation of the bootstrap sample) of 8.59 Mgha^{-1} (Figure SI6a and b). The model for wetlands is also unbiased (-1.19 Mgha^{-1}) with smaller RMSE (19.28 Mgha^{-1}) and standard error (2.76 Mgha^{-1}, (Figure SI6c and d). We calculated the weighted arithmetic average of the RMSE observed for the 40 plots over drylands and the 16 plots over wetlands to estimate the overall RMSE of the lidar-AGB model that is of 49.94 Mgha^{-1}. We found a strong relationship for both models with R^2=0.81 for drylands and R^2=0.79 for wetlands. R^2 is unbiased for both models with very low standard errors (0.06 and 0.07, respectively)
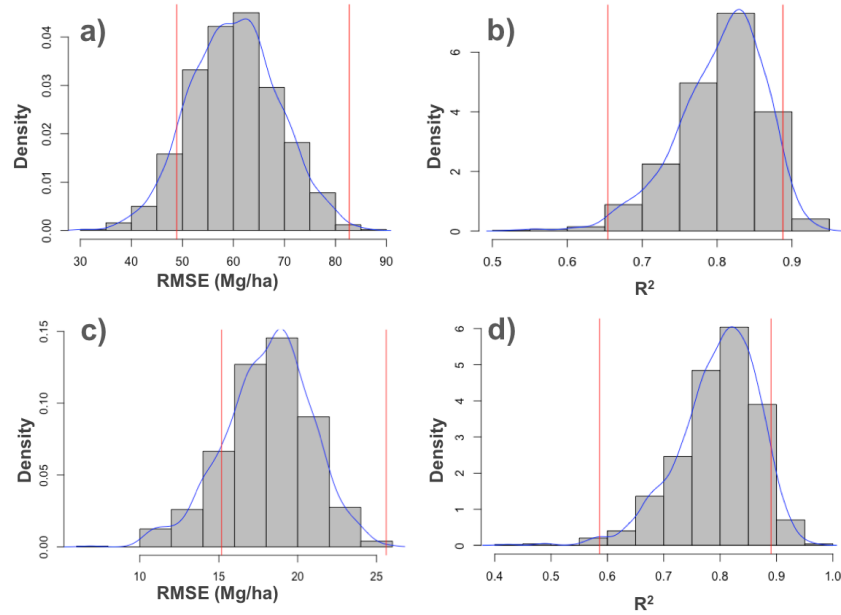
Figure SI6. Density histogram with a kernel density line (blue line) and the 95 % confidence interval (red lines) regarding the bootstrapping (n=1000) approach for the **a)** RMSE and **b)** Pearson coefficient of the lidar-AGB model of drylands forests. In **c)** and **d)** we show the corresponding results for wetlands. The 95% confidence interval values correspond to 48.46 Mgha^{-1} and 78.39 Mgha^{-1} in **a)**, 0.61 and 0.85 in **b)**, 1.18 Mgha^{-1} and 25.60 Mgha^{-1} in **c)** and, 0.59 and 0.89 in **d)**.

6.2 RF model

The uncertainty of the RF model is estimated using a bootstrapped (15 times) leave-30%-out cross validation approach where 70% of the 54 lidar scenes are used for prediction purposes and the remaining ones for testing. Results for each iteration and corresponding averaged values are show in Table SI5. RMSE ranges from 92.1 Mgha^{-1} to 123.8 Mgha^{-1} with an average of 101.9 Mgha^{-1}. The relationship between observed and predicted AGB is moderately strong (R^2=0.5) and unbiased (0.49 Mgha^{-1}). Figure SI7a is a scatter plot corresponding to one iteration of the cross-validation method where each point corresponds to 1 ha pixel from the lidar-sampling that has been kept out from the RF algorithm to compare with the RF map. Table SI5 and Figure SI7a both show that local errors can be significantly high (mainly for the densest areas of our region of interest) but the fact that estimates are unbiased ensures that we can estimate averaged AGB values over large areas with confidence (Section 3.6). The reasons for these local errors are discussed in Section 4.2 and mainly concern the inability of the satellite remote sensing layers to describe the forest structure variability (mainly over dense tropical areas) and the loss of interpretably of the RF algorithm towards the tails of the AGB distribution. In fact, our results compare with a recent study over the Democratic Republic of Congo for a similar analysis for a wall-to-wall map with 1 ha grid cells the Democratic Republic of Congo (DRC, ~ 100 Mgha^{-1}, Xu *et al.,* 2017).

Iteration	R^2	RMSE (Mgha^{-1})	Bias (Mgha^{-1})
1	0.61	92.1	-6.3
2	0.49	114.2	2.4
3	0.54	109.1	1.3
4	0.63	87.9	-2.0
5	0.51	123.8	-3.0
6	0.63	89.7	4.5
7	0.62	98.1	9.5
8	0.51	104.7	14.8
9	0.61	103.9	26.3
10	0.62	90.7	-12.0
11	0.67	87.3	-18.4
12	0.51	113.2	-12.0
13	0.49	105.5	-5.7
14	0.61	92.2	6.2
15	0.56	115.9	1.7
Total	**0.5±0.06**	**101.9±11.3**	**0.49 ±10.9**

6.3 Uncertainty map
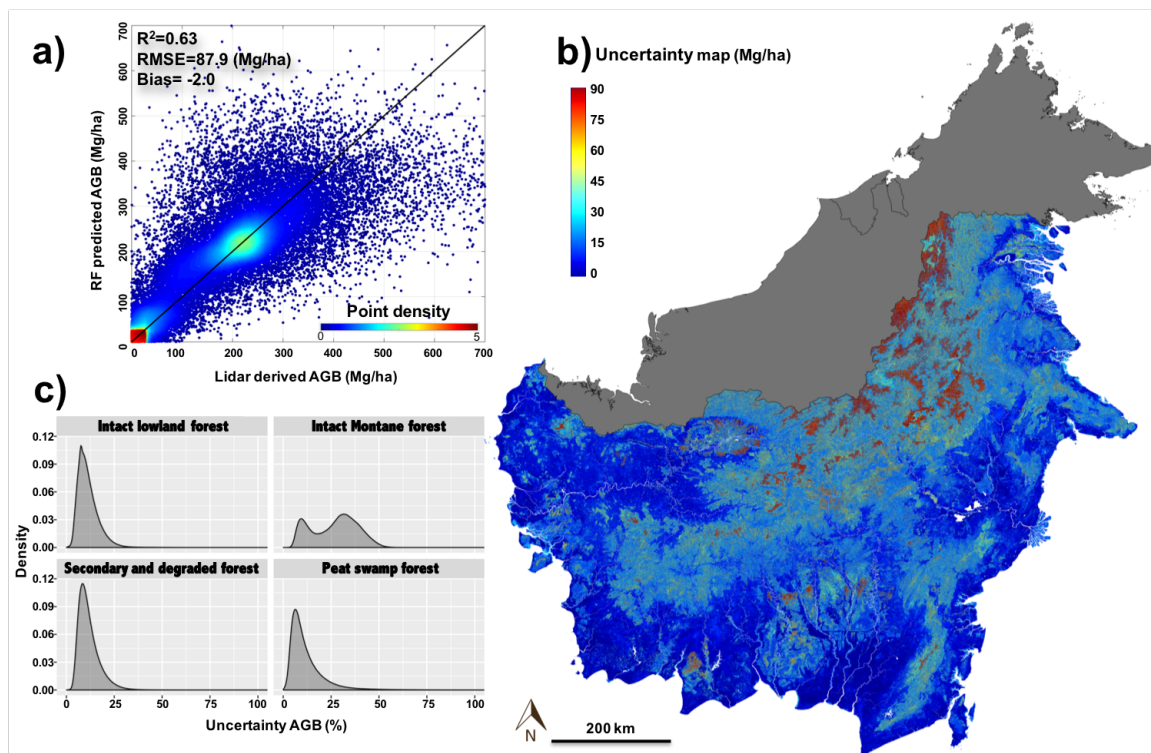
Figure SI7. Uncertainty analysis with a) an example of a scatter plot representing the errors issued from the bootstrapping leave-30%-out cross validation method (iteration 4 in Table SI5), b) pixel level uncertainty map calculated as explained in SI6.3 and, c) the uncertainty distribution summarized for the target land cover classes in terms of percentage with respect to the corresponding mean AGB.

Unlikely other methods (e.g. maximum entropy), RF does not provide a pixel-based uncertainty for regression problems. The cross-validation approach described in SI6.2 provides a series of 15 different AGB density maps that are combined to calculate a pixel-level uncertainty product by calculating the standard deviation of the corresponding pixels Figure SI7b. Higher standard deviation values for a given area indicate that AGB is estimated with a higher uncertainty means that results are not stable and that they depend on the random subsampling of the training data. Therefore, our uncertainty map represents the stability of our model predictions spatially across gradients of forest structure and surface topography. It shows that errors increase with elevation that is directly correlated with rougher topography.

The overall average pixel-level uncertainty is of 23.6 ± 20.5 Mgha^{-1} (SI6.4) and it varies within a narrow range (26.6-30.0 Mgha^{-1}) depending on the target land cover class. The exception is the intact montane forest whose uncertainty is significantly higher (73.4 ± 32.3 Mgha^{-1}, Table SI6) and it is the only forest type that has uncertainty higher than 25% regarding the mean AGB (Figure SI7c). The error has a bi-modal distribution that seems to follows the variance of topographic slope. The large uncertainty associated with montane AGB is mainly due to the reduced ability of remote sensing products (namely satellite radar) to properly characterize forest structure over rough terrain. Finally, we observed higher pixel-level uncertainty over highly dense primary peat swamp tall pole forests close to the borders with drylands forests (e.g. lake Sentarum and north part of Sebangau National Park). However, they are not significant in terms of percentage concerning the peat swamp forest (Figure SI7c). Pixel-level uncertainty is also relatively low in terms of percentage regarding intact and secondary and degraded forest.

Table SI6. Average uncertainty of the AGB density map by land cover class derived from the IMF land cover map (SNI, 2016) $\pm$ standard deviation of the distribution.

Land Cover	Uncertainty ± S.D. (Mgha^{-1})
Intact Lowland Forest	30.0 ± 14.4
Intact Montane Forest	73.4 ± 32.3
Secondary and Degraded Forest	26.6 ± 12.5
Peat Swamp Forest	26.9 ± 21.1
Swamp Scrublands	9.12 ± 8.9
Tree Plantations	6.2 ± 4.6
Scrublands	12.2 ± 10.1
Total	**23.6 ± 20.5**

6.4 Uncertainty in estimating average AGB

We assessed the uncertainty associated with the estimation of the average AGB for a given region of interest (ROI) such as a forest type or administrative region by propagating local- and pixel-level errors to the entire ROI taking into account spatial autocorrelation and covariance of errors. By definition, the error propagation and overall uncertainty assessment must include errors associated with 1) the measurements in the field inventory (e.g. error in measuring diameter at breast height, D), 2) remote sensing estimates (e.g. lidar forest height), 3) field-based estimation of AGB from pan-tropical allometry to

convert D into AGB, 4) the lidar-AGB parametric model uncertainty, 5) RF non-parametric model uncertainty for spatial mapping, and 6) the errors associated with the land cover classification. However, we only included uncertainty associated with the lidar-AGB model and the geospatial RF modeling because we had no data to evaluate the remaining sources of errors or they are considered negligible compared to the aforementioned (Garcia et al., 2017a). For instance, the land cover mapping misclassification should be confined to the classes boundaries and areas with recent events (e.g. deforestation) and are expected to have little impact on the region-level statistics.

We follow (Chen et al., 2015) to assess the uncertainty associated with the estimation of the mean AGB for a given ROI:

$$\sigma_{AGB}^2(ROI) = \sigma_{lm}^2(ROI) + \sigma_{rf}^2(ROI) \qquad \text{SI2}$$

where $\sigma_{lm}^2(ROI)$ is the error associated with the lidar-AGB model and $\sigma_{rf}^2(ROI)$ the error of the RF model. As far as the Lidar-AGB model error is concerned, we define:

$$\sigma_{lm}^2(ROI) = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} cov(\sigma_{l,i}, \sigma_{l,j})\right) \qquad \text{SI3}$$

where N is the total number of grid cells within a given ROI and, (i, j) is the index for the cell location. Equation SI2 can be reformulated as

$$\sigma_{lm}^2(ROI) = \sum_{p=1}^{m}\quad\sum_{q=1}^{m}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial f}{\partial \phi_p}\right) cov(\phi_p, \phi_q)\left(\frac{1}{N}\sum_{j=1}^{N}\frac{\partial f}{\partial \phi_q}\right) = \sum_{p=1}^{m}\sum_{q=1}^{m}\left(\bar{g}_p cov(\phi_p, \phi_q)\bar{g}_q\right) \qquad \text{SI4}$$

Where $\bar{g}_p = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial f}{\partial \phi_p}$ is the mean of the first derivative with respect to the coefficient ϕ_p of the allometric model and m is the total number of coefficients in the lidar-AGB model (Eq. 1). Note that m equals 2 for each lidar-AGB model that have been developed for drylands and wetlands (Figure 3, refer to SI6.1 for details on the lidar-AGB modelling and plot-level uncertainty).

The uncertainty associated with the RF predictions given by the second term in Equation SI2 can be modelled using the following covariance:

$$\sigma_{rf}^2(ROI) = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma_{r,i}^2 + \sum_{i=1}^{N}\quad\sum_{i\neq j}\rho_{i,j}\sigma_{r,i}\sigma_{r,j}\right) \qquad \text{SI5}$$

Where σ_r is the pixel-level uncertainty from the RF modeling (refer to SI6.3 for details on the uncertainty map) and $\rho_{i.j}$ is the correlation coefficient between cells i and j,

which can be estimated using the following variogram under the assumption that spatial autocorrelation only changes with distance (h):

$$\gamma(h) = \tfrac{1}{2}E\left[\left(y_{x_i} - y_{x_j}\right)^2\right] = [C(0) - C(h)], \text{ with } \left\|(x_i - x_j)\right\| = h \qquad\qquad \text{SI6}$$

where $C(h)$ is the covariogram depending on the distance h. The resulting variogram is show in Figure SI8. Results on the uncertainty in estimating average AGB using Equation SI2 at the land cover class-level, region-level and Kalimantan-level are reported in Table 1, Table 2, Table SI3 and, Table SI4. Note that the uncertainty regarding the lidar-based estimates show in Table 2 are calculated using only $\sigma_{lm}^2(ROI)$.
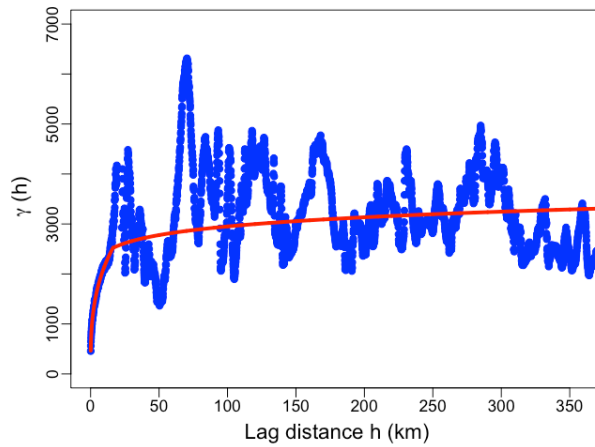
Figure SI8. Variogram of the prediction residual showing spatial autocorrelation changing with the lag distance. The range, nugget and sill equal 16.5 km, 21.8 Mgha[-1] and 47.3 Mgha[-1].

6.5 Error assessment using field inventory data

We compare the wall-to-wall RF map directly to field derived AGB estimates as an additional assessment of the mapping error. We consider the external testing datasets (Sindo Lumber, Timberdana, Rodamas and USFS, Table SI1 and Table SI2) as well as the field plots that have been used to calibrate the lidar AGB estimator (Malinau, Sumalindo IV and Mawas). The inventory sites Mawas and USFS are located over peat swamp forests whereas the remaining ones correspond to dryland forests. However, there is a disagreement between the AGB map resolution (1 ha) and the size of the field plots, which range from 0.1ha to 0.25ha (Table SI1). We aggregated the field plots to 1 ha (by aggregating the four nearest 0.25 ha or 0.1 ha plots, its barycenter being the location of the new 1 ha plot) in order to properly compare with the AGB map. We calculated the average AGB for the plots that covered more than one 1 ha pixel. The result is show in Figure 5 (R^2= 0.73 and RMSE=77.8 Mgha[-1]) and it is relatively lower compared to the results found in the Democratic Republic of Congo

for a similar analysis at 1-ha grid cells (R^2= 0.5 and RMSE=89.72 Mgha^{-1}, Xu *et al.,* 2017).

References

Asner G and Mascaro J 2014 Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric *Remote Sens. Environ.* **140** 614–24 Online: http://dx.doi.org/10.1016/j.rse.2013.09.023

ASPRS 2015 *LAS SPECIFICATION VERSION 1.4 – R13* Online: http://www.asprs.org/wp-content/uploads/2010/12/LAS_1_4_r13.pdf

Baccini a., Goetz S J, Walker W S, Laporte N T, Sun M, Sulla-Menashe D, Hackler J, Beck P S a., Dubayah R, Friedl M a., Samanta S and Houghton R a. 2012 Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps *Nat. Clim. Chang.* **2** 182–5 Online: http://dx.doi.org/10.1038/nclimate1354

Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32

Canty A and Ripley B 2017 boot: Bootstrap R (S-Plus) Functions 117 Online: https://cran.r-project.org/web/packages/boot/index.html

Chave J, Réjou-Méchain M, Búrquez A, Chidumayo E, Colgan M, Delitti W, Duque A, Eid T, Fearnside P, Goodman R, Henry M, Martínez-Yrízar A, Mugasha W, Muller-Landau H, Mencuccini M, Nelson B, Ngomanda A, Nogueira E, Ortiz-Malavassi E, Pélissier R, Ploton P, Ryan C, Saldarriaga J and Vieilledent G 2014 Improved allometric models to estimate the aboveground biomass of tropical trees *Glob. Chang. Biol.* **20** 3177–90 Online: http://dx.doi.org/10.1111/gcb.12629

Chen Q, Vaglio Laurin G and Valentini R 2015 Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels *Remote Sens. Environ.* **160** 134–43 Online: http://www.scopus.com/inward/record.url?eid=2-s2.0-84923286985&partnerID=tZOtx3y1

Ferraz A, Mallet C and Chehata N 2016 Large-scale road detection in forested mountainous areas using airborne topographic lidar data *ISPRS J. Photogramm. Remote Sens.* **112** 23–36

Galudra G, Van Noordwijk M, Suyanto S, Sardi I, Pradhan U and Catacutan D 2011 Hot spots of confusion: contested policies and competing carbon claims in the peatlands of Central Kalimantan, Indonesia *Int. For. Rev.* **13** 431–41 Online: http://openurl.ingenta.com/content/xref?genre=article&issn=1465-5489&volume=13&issue=4&spage=431

Garcia M, Saatchi S, Casas A, Koltunov A, Ustin S, Ramirez C, Garcia-Gutierrez J and An 2017a Quantifying biomass consumption and carbon release from the California Rim fire by integrating airborne LiDAR and Landsat OLI data *J. Geophys. Res. Biogeosciences* **122** 340–53

Garcia M, Saatchi S, Ferraz A, Silva C A, Ustin S, Koltunov A and Balzter H 2017b

Impact of data model and point density on aboveground forest biomass estimation from airborne LiDAR *Carbon Balance Manag.* **12**

GFW Indonesia Ministry of Forestry, Greenpeace, and WRI. "Indonesia logging concessions." Accessed through Global Forest Watch on 17/11/2017. www.globalforestwatch.org.

Di Gregorio A and Jansen L 2000 *Land Cover Classification System (LCCS): Classification Concepts and User Manual* (Rome, Italy: Food and Agriculture Organization of the United Nations) Online: http://www.fao.org/docrep/003/x0596e/x0596e00.HTM

Guo L, Chehata N, Mallet C and Boukir S 2011 Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests *ISPRS J. Photogramm. Remote Sens.* **66** 56–66 Online: http://dx.doi.org/10.1016/j.isprsjprs.2010.08.007

Hansen M C, Potapov P V., Moore R, Hancher M, Turubanova S A, 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 (80-.).* **342** 850–3 Online: http://www.sciencemag.org/cgi/doi/10.1126/science.1244693

Hoekman D H, Vissers M A M and Wielaard N 2010 PALSAR wide-area mapping of Borneo: Methodology and map validation *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **3** 605–17

Mascaro J, Asner G P, Knapp D E, Kennedy-Bowdoin T, Martin R E, Anderson C, Higgins M and Chadwick K D 2014 A tale of two "Forests": Random Forest machine learning aids tropical Forest carbon mapping *PLoS One* **9** 12–6

Meledy, L., Hagen, S., Sullivan, F.B., Pearson, T., Walker, S.M., Ellis, P., Kustiyo Sambodo, K.A., Roswintiarti, O., Hanson, M., Klassen, A.W., Palace, M.W., Braswell, B.H., Delgado, G.M., Saatchi, S.S., Ferraz, A., 2017. CMS: LiDAR Data for Forested Sites on Borneo Island, Kalimantan, Indonesia, 2014. doi:10.3334/ornldaac/1518

Meyer, V., Saatchi, S., Clark, D., Keller, M., Vincent, G., Ferraz, A., Espírito-Santo, F., d'Oliveira, M. V.N., Kaki, D., Chave, J., 2018. Canopy area of large trees explains aboveground biomass variations across nine neotropical forest landscapes. Biogeosciences, 15, 3377-3390. doi.org/10.5194/bg-15-3377-2018, 2018.

Miettinen J, Shi C., and Liew S C 2016 Land cover distribution in the peatlands of Peninsular Malaysia, Sumatra and Borneo in 2015 with changes since 1990 *Glob. Ecol. Conserv.* **6** 67–78 Online: http://dx.doi.org/10.1016/j.gecco.2016.02.004

Saatchi S S, Harris N L, Brown S, Lefsky M, Mitchard E T A, Salas W, Zutta B R, Buermann W, Lewis S L, Hagen S, Petrova S, White L, Silman M and Morel A 2011 Benchmark map of forest carbon stocks in tropical regions across three continents *Proc. Natl. Acad. Sci.* **108** 9899–904 Online: http://www.pnas.org/content/108/24/9899.abstract

Sephton T and Wishart A 2013 TOPOLEV and C-PARAS BT - Distributed Space Missions for Earth System Monitoring ed M D'Errico (New York, NY: Springer New York) pp 473–508 Online: https://doi.org/10.1007/978-1-4614-4541-8_16

SNI 2016 Klasifikasi Penutup Lahan (Land Cover Classification) *Minist. For. Jakarta* Online: http://nfms.dephut.go.id/ipsdh/index.php

Theobald D M, Stevens D L, White D, Urquhart N S, Olsen A R and Norman J B 2007 Using GIS to generate spatially balanced random survey designs for natural resource applications *Environ. Manage.* **40** 134–46

Tittmann P, Saatchi S and Sharma B 2015 VCS: Tool for measuring aboveground live forest biomass using remote sensing 1–32

UNFCCC 2002 *A Guide to the Climate Change Convention and its Kyoto Protocol* (Bonn)

Xu L, Saatchi S S, Shapiro A, Meyer V, Ferraz A, Yang Y, Bastin J-F, Banks N, Boeckx P, Verbeeck H, Lewis S L, Muanza E T, Bongwele E, Kayembe F, Mbenza D, Kalau La, Mukendi F, Ilunga F and Ebuta D 2017 Spatial distribution of carbon stored in forests of the Democratic Republic of Congo *Sci. Rep.* 1–12

Zhang K, Chen S C, Whitman D, Shyu M L, Yan J and Zhang C 2003 A progressive morphological filter for removing nonground measurements from airborne LIDAR data *IEEE Trans. Geosci. Remote Sens.* **41** 872–82