

## Supplementary information

### Correlation computation

To get a quantitative assessment of the correlation and the dispersion of L-VOD versus the evaluation datasets, three correlation coefficients were computed. First, the Pearson correlation coefficient  $R$  of two variables  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$  was computed as:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{S1})$$

where,  $\bar{x}$  and  $\bar{y}$  are the means of each variable.  $R$  is a measure of the linear correlation between two variables. If the relationship linking these variables is linear with no dispersion,  $R$  equals 1 (both variables increase together) or -1 (one variable increases when the other decreases).

However, the relationships between L-VOD and the evaluation data are not expected to be linear in most of the cases. Therefore, two rank correlations were also computed to quantify monotonic relationships whether linear or not. The Spearman's correlation coefficient  $\rho$  is the Pearson correlation coefficient  $R$  computed on the rank of the two variables instead of the variables themselves. If there are no repeated data values, a perfect Spearman correlation of +1/-1 occurs when each of the variables is a perfect monotonic function of the other. In addition, Kendall's rank correlation was also computed. Kendall's correlation coefficient  $\tau$  is given by:

$$\tau = \frac{n_{\text{concordant}} - n_{\text{discordant}}}{n(n-1)/2} \quad (\text{S2})$$

where  $n_{\text{concordant}}$  and  $n_{\text{discordant}}$  are the number of concordant and discordant pairs, respectively. Given a pair of observations  $(x_i, y_i)$  and  $(x_j, y_j)$ , they are said to be concordant if  $y_j > y_i$  for  $x_j > x_i$  or  $y_j < y_i$  for  $x_j < x_i$ . Otherwise, the pair is said to be discordant. The denominator is the total number of pair combinations, so  $\tau$  is in the range [-1,1].

### Non-linear fits to the AGB versus L-VOD relationship

The relationships linking L-VOD to the evaluation data for different biomes were fitted using linear fits. In addition, fits to the global relationships linking AGB and L-VOD were computed following the approach used by Liu et al. (2015). The SMOS-IC

L-VOD data was binned in 0.05-width bins. For each L-VOD bin, the 5th and 95th percentiles and the mean of the AGB distribution were computed, obtaining three AGB curves as a function of L-VOD. The three curves were fitted with Liu's function, with a logistic function or with a generalized logistic function, obtaining results of the same quality. Figure S3 shows the fitted curves and Table S2 presents the parameters of the fits obtained using a logistic function:

$$5 \quad AGB = \frac{a}{1 + e^{-b(VOD-c)}} + d \quad (S3)$$

where  $a, b, c, d$  are four best-fit parameters. The fitted curves give AGB in Mg/h units as a function of L-VOD, which is dimensionless quantity. Therefore the units of  $a$  and  $d$  are Mg/h and  $b$  and  $c$  are dimensionless quantities. Table S2 also gives the values of the best-fit parameters and the correlation coefficients between the observed and fitted L-VOD data.

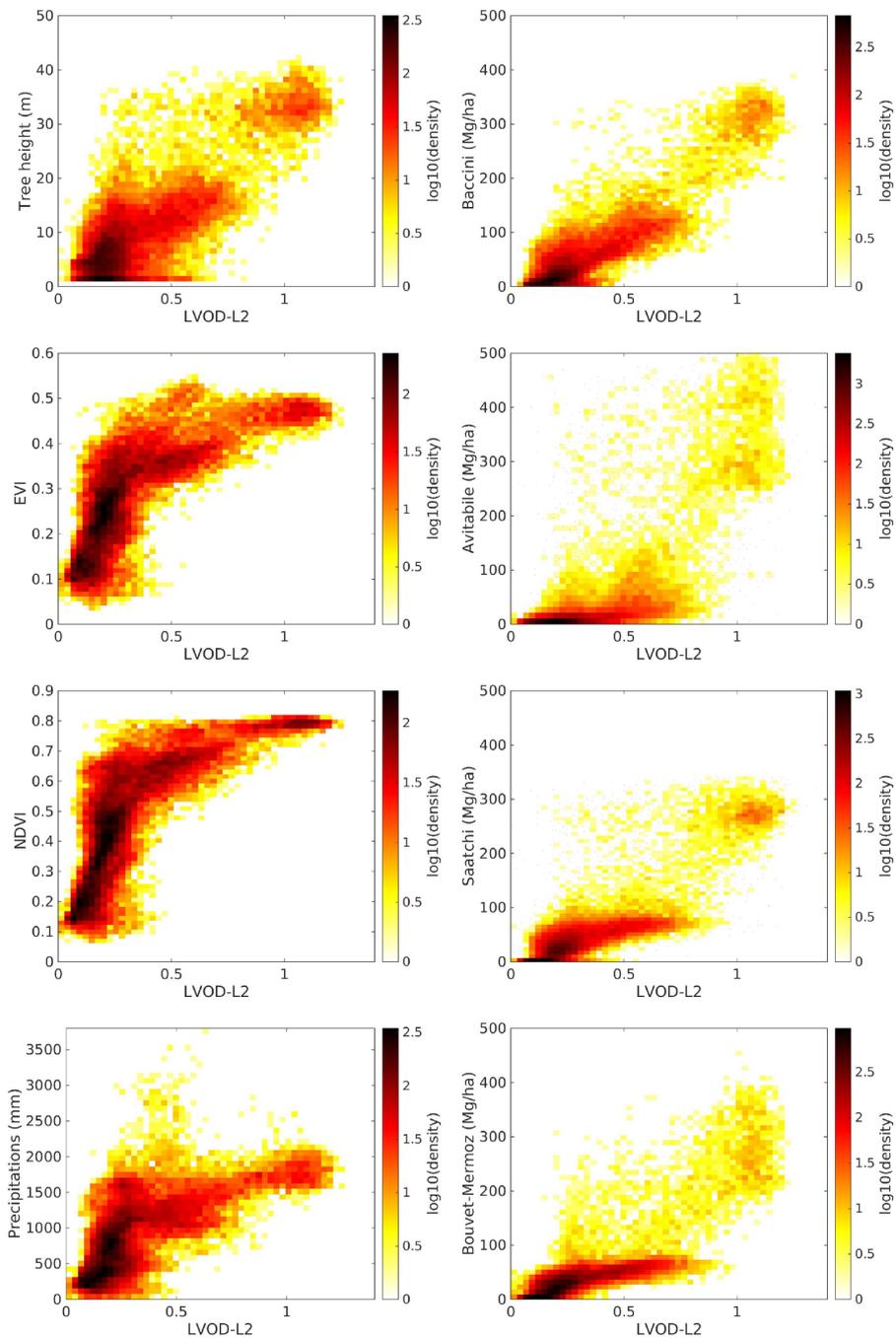
To compare the performance of L-VOD to estimate AGB with respect to other vegetation indices, scatter plots similar to those of Fig. 2 were computed using Saatchi's AGB with respect to MODIS NDVI and EVI (Fig. S4). There is a close-to-linear relationship for AGB lower than  $\sim 90$  Mg/h and EVI or NDVI lower than 0.4 and 0.7, respectively. However, in contrast to L-VOD, the relationship saturates for higher AGB values and both EVI and NDVI show a very low sensitivity to AGB with increments of 90 to 300 Mg/h in the 0.5-0.6 and 0.7-0.8 ranges for EVI and NDVI, respectively. This is expected as the visible/infra-red indices are sensible to the greenness of the canopy, which is not closely related to the total AGB in densely vegetated regions.

### L-VOD versus AGB and tree height for different biomes

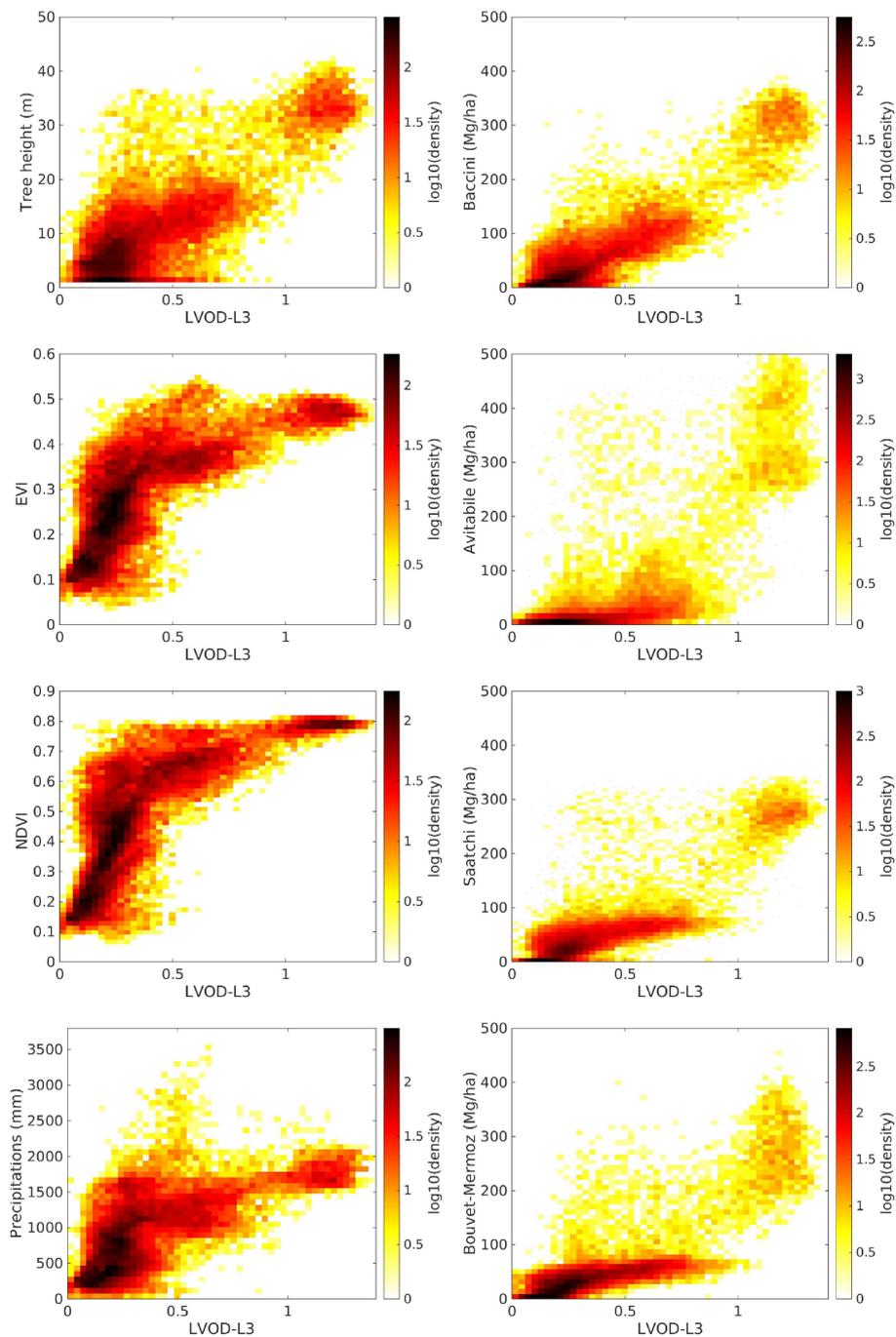
Figure S5 shows the distribution of the two biomes groups used to compute Fig. 3. In addition, Fig. S6 shows the L-VOD versus the AGB and tree height data sets using the more specific land cover classes shown in Fig. S7. Most of the relationships are close to linear with  $R \sim 0.7$  or higher, and  $\rho$  values similar to  $R$ . For all land cover classes, the correlations with respect to Avitabile AGB are the lowest of all the AGB data sets, except for the Mermoz AGB for evergreen broadleaf rainforest. The highest correlations for shrublands, croplands, natural vegetation and grasslands and savannahs were found with Saatchi's and Bouvet's AGB ( $R = 0.73-0.79$  and  $\rho = 0.73-0.78$ ), while the highest correlations were found with Baccini's AGB for woody savannah. For dense evergreen broadleaf forest the highest correlations were found with Baccini's and Saatchi's AGB. For the latter vegetation type, the values of the correlation coefficient values are comparable to those obtained for the other land cover classes but the slope of the regression lines are significantly higher (by factors of 1.3-1.9) than those obtained for shrublands, croplands, grasslands, and savannahs, which are rather similar for Baccini, Saatchi and Bouvet-Mermoz data sets ( $\sim 110-140$  Mg/h). Once again, the L-VOD relationships obtained between L-VOD with respect to and Avitabile AGB are very different to those found with the other AGB data sets, with slopes changing by a factor of 8 from that computed for savannahs to that obtained for evergreen broadleaf forest, leading to the highly non-linear global relationship discussed in Sect 4. Baccini's AGB data set shows a high dynamical range for low AGBs, as AGB is as high as  $\sim 150$  Mg/h for IC L-VOD  $\sim 0.6$ , while for the other data sets, the maximum AGB for this L-VOD value is less than 100 Mg/h.

**Table S1.** Main characteristics of the three SMOS L-VOD products used in this study.

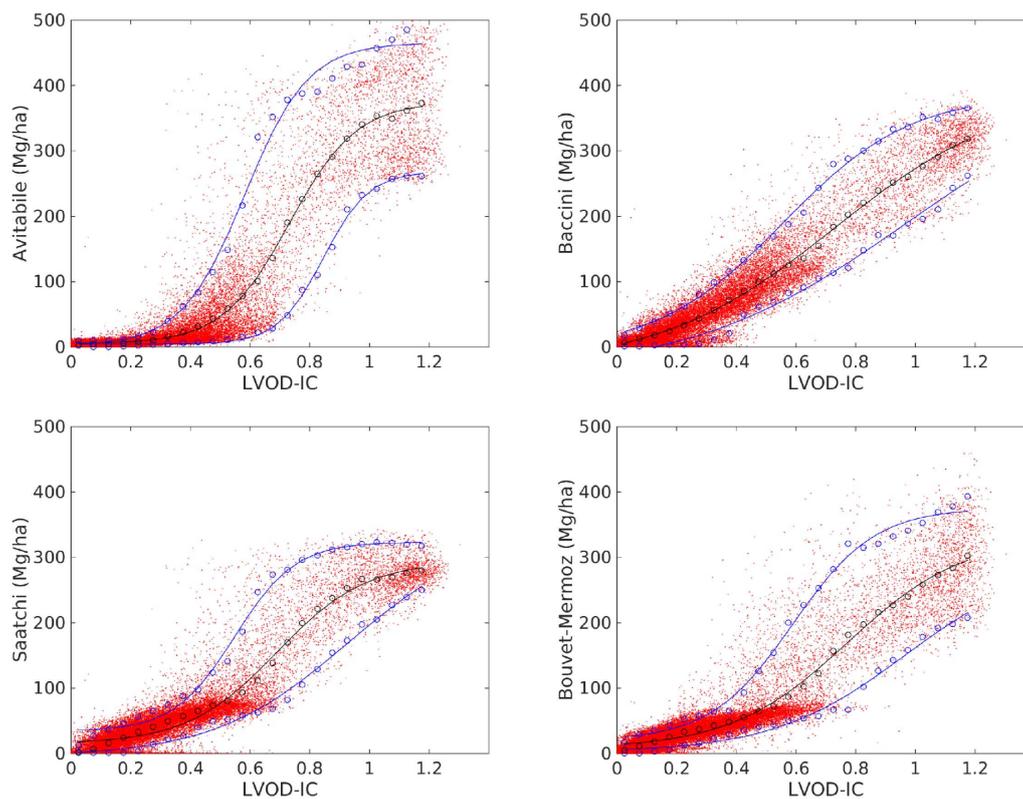
	ESA Level 2	CATDS Level 3	INRA-CESBIO
Reference	Kerr et al. (2012)	Al Bitar et al. (2017)	Fernandez-Moran et al. (2017)
Version used	620	300	100
Soil texture	Ecoclimap	Ecoclimap	Ecoclimap
Land cover	Ecoclimap	Ecoclimap	IGBP
Soil temperature	ECMWF	ECMWF	ECMWF
Forward model	L-MEB (Wigneron et al., 2007)	L-MEB (Wigneron et al., 2007)	L-MEB (Wigneron et al., 2007)
Multi-orbit	no	yes, three orbits, L-VOD assumed to be correlated	no
L-VOD first guess	Computed from Ecoclimap LAI	Computed from Ecoclimap LAI	First inversion using a constant value of 0.2 and a local average of previous IC retrievals in a second step
SM first guess	ECMWF	ECMWF	$0.2 \text{ m}^3/\text{m}^{-3}$
Footprints with inhomogeneous land cover	SM and L-VOD retrieval only for major fraction. Contribution from minor fraction using ECMWF SM and Ecoclimap LAI	SM and L-VOD retrieval only for major fraction. Contribution from minor fraction using ECMWF SM and Ecoclimap LAI	SM and L-VOD retrieval for the whole footprint assumed to be homogeneous
Grid	ISEA	EASEv2	EASEv2
Sampling	15 km	25 km	25 km



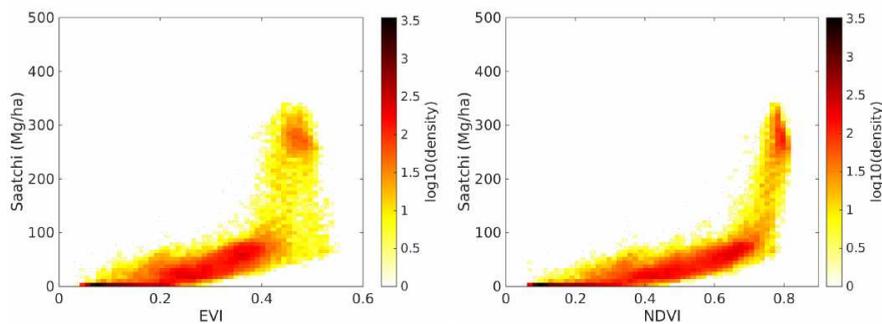
**Figure S1.** Density scatter plots of the 2011 annual mean of SMOS L2 L-VOD respect to (from top to bottom and from left to right): tree height, EVI, NDVI, cumulated precipitation, Baccini et al. (2012), Avitabile et al. (2016), Saatchi et al. (2011) and Bouvet-Mermoz AGB datasets.



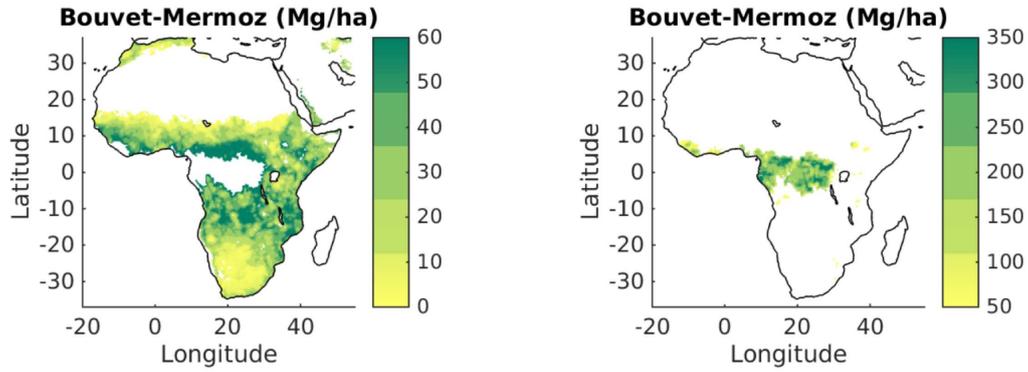
**Figure S2.** Density scatter plots of the 2011 annual mean of SMOS L3 L-VOD respect to (from top to bottom and from left to right): tree height, EVI, NDVI, cumulated precipitation, Baccini et al. (2012), Avitabile et al. (2016), Saatchi et al. (2011) and Bouvet-Mermoz AGB datasets.



**Figure S3.** AGB versus L-VOD scatter plots of Fig. 2 but plotted as point scatter plots. In addition, on the right-hand panels, the 5th and 95th percentiles of the AGB distribution in bins of L-VOD are displayed as blue circles while the mean is displayed as black circles. Solid blue and black lines are the fits obtained using a logistic function (Eq. S3) with the parameters given in Table S2 for the 5th and 95th percentiles and the mean curves.



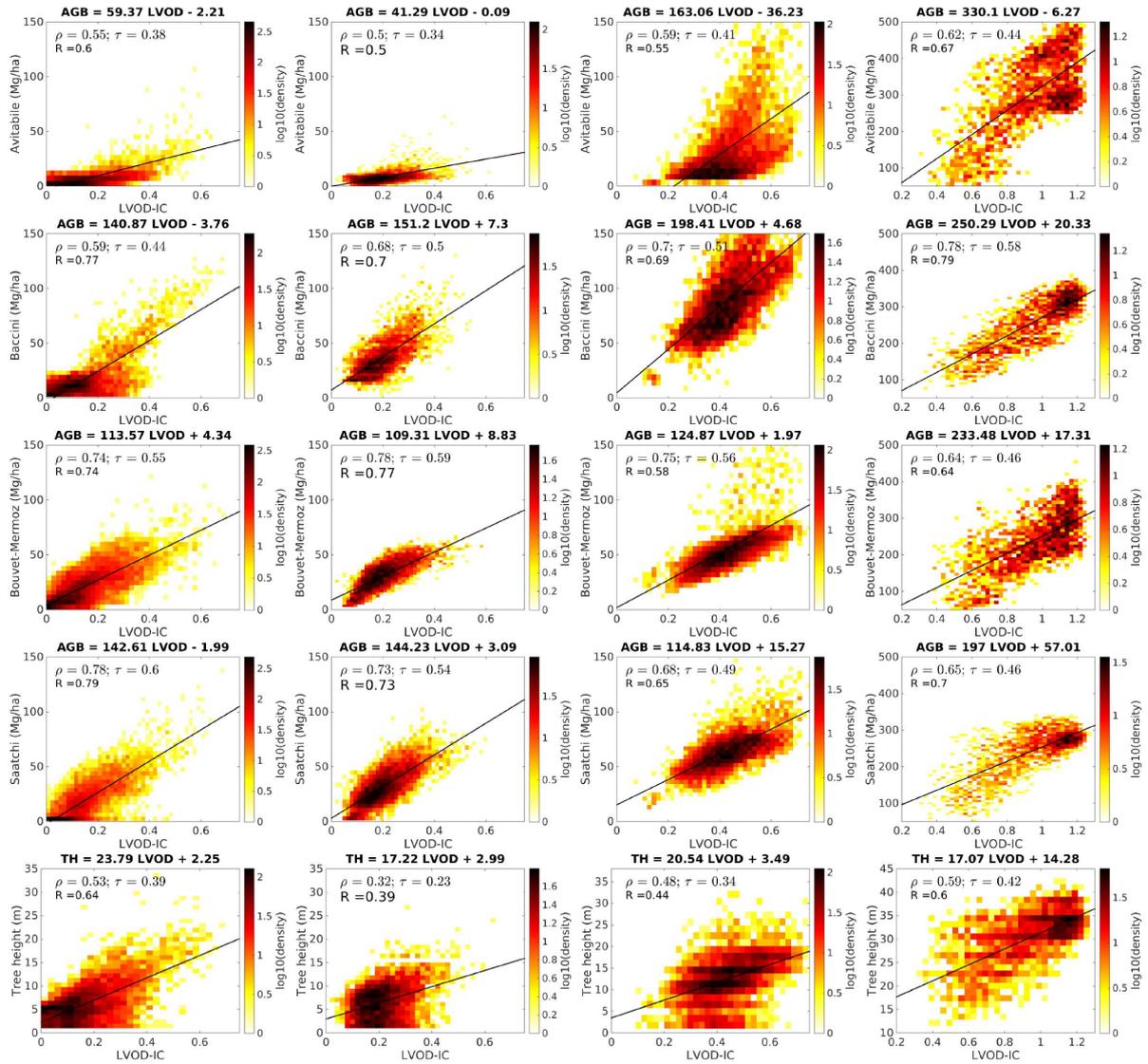
**Figure S4.** Scatter plots of MODIS NDVI and EVI with respect to Saatchi et al. (2011) AGB.



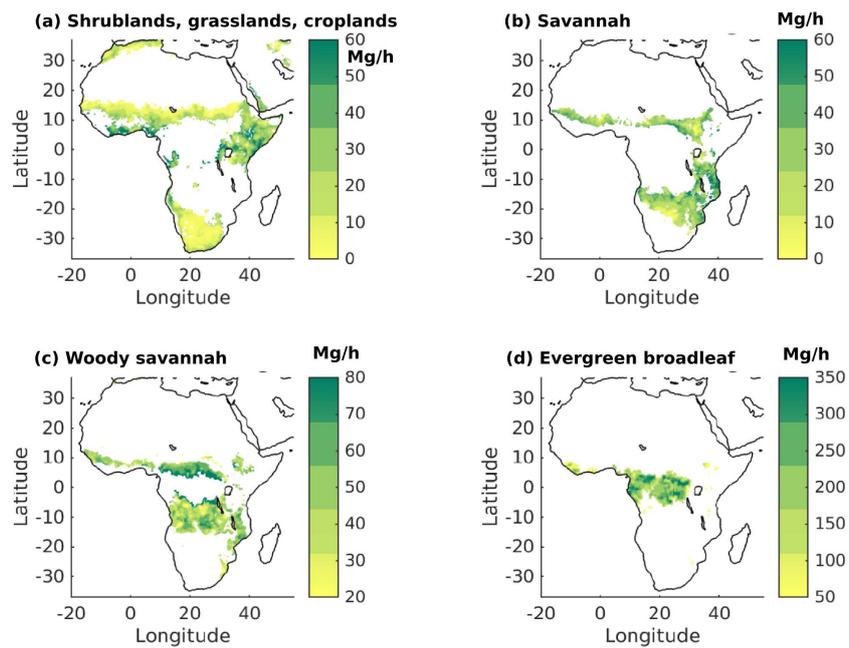
**Figure S5.** Bouvet-Mermoz data set showing the spatial distribution of the classes used to compute Fig. 3.

**Table S2.** Parameters of the fits of the AGB vs IC L-VOD of relationship of Fig. S3 using a logistic function (Eq. S3).

AGB	line	$a$ [Mg/h]	$b$ [-]	$c$ [-]	$d$ [Mg/h]	$R^2$
Avitabile	05th	264.367	13.115	0.846	4.351	0.998
Avitabile	Mean	369.890	8.921	0.732	5.158	0.999
Avitabile	95th	463.091	9.466	0.583	2.135	0.990
Saatchi	05th	345.590	4.458	0.926	-4.387	0.993
Saatchi	Mean	280.159	6.680	0.689	14.794	0.993
Saatchi	95th	289.762	9.857	0.548	33.859	0.993
Baccini	05th	455.774	2.785	0.964	-40.357	0.990
Baccini	Mean	422.744	3.400	0.729	-29.252	0.999
Baccini	95th	393.863	4.685	0.558	-6.444	0.997
Bouvet-Mermoz	05th	296.709	4.511	0.966	2.129	0.987
Bouvet-Mermoz	Mean	325.043	5.116	0.774	7.651	0.996
Bouvet-Mermoz	95th	355.989	7.267	0.589	19.731	0.994



**Figure S6.** SMOS IC L-VOD relationships to the AGB and tree height evaluation datasets for different land cover classes: From left to right: (i) Shrublands, croplands, natural vegetation and grasslands, (ii) Savannah (ii) Woody savannah, (iv) Evergreen broadleaf



**Figure S7.** Spatial distribution of the Bouvet-Merzoz AGB for the land cover classes used to compute the scatter plots of Fig. S6. **(a)** Shrublands, croplands, natural vegetation and grasslands, **(b)** Savannah **(c)** Woody savannah, **(d)** Evergreen broadleaf