

Supplementary information

Figure S1. Machine learning workflow for the generation of global wood density (WD) maps. The predictor covariates, highlighted in light blue highlighted, include climate climatology (CLM), climate extreme indexes (EXM), topography (TOPO), soil properties (SOIL), vegetation characteristics (VEG), and land cover types (LC). These covariates are fed into four machine learning models highlighted in red, which include light gradient boosting model (LGBM), Scikit-Learn Random Forest (RF), LightGBM Random Forest (LGBM-RF), and extreme gradient boosting model (XGBoost). In addition, eight cross-validation strategies, highlighted in purple, including random 5-fold (Random), spatial-blocked 10-fold (Spatial-blocked), CCI LC classification (CCI), FAO ecological zone (FAO), two Köppen climate classifications (Köppen), 2-deg latitudinal and 5-deg longitudinal zones (Lat/Lon) are used to test the extrapolation capacity of machine learning models.

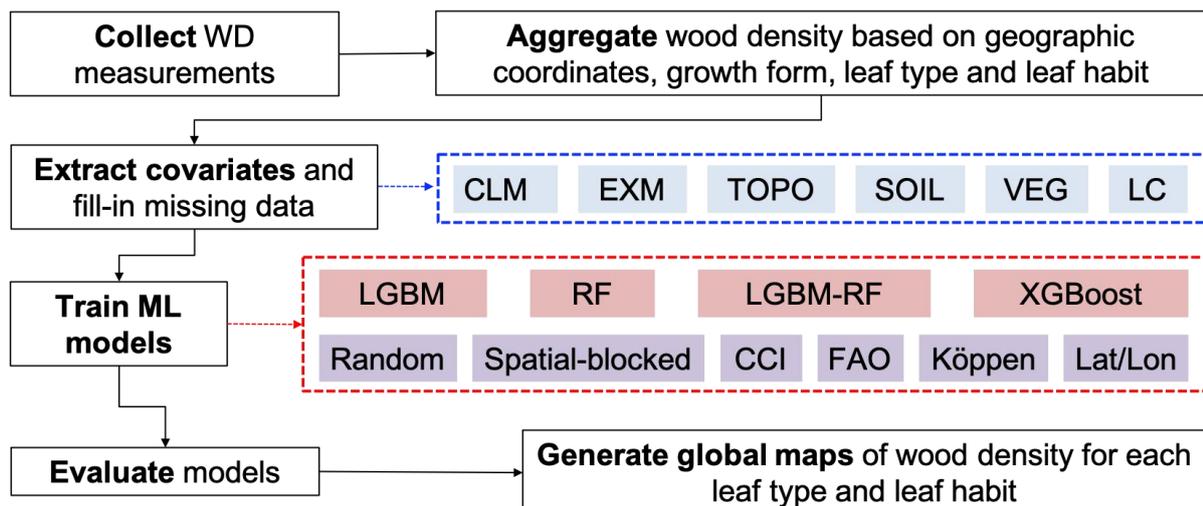


Figure S2. (a) Model performance: R^2 and root mean squared error (RMSE) for the prediction of training data, cross-validation (CV) data and test data using four distinct machine learning models, i.e., LightGBM, LightGBM-RF, Random Forest and XGBoost. Displayed are results under eight distinct cross-validation methods. (b) R^2 and RMSE for the eight different cross-validation methods: 1) random 5-fold, 2) spatial blocked 10-fold, 3) latitude 2-degree, 4) longitude 5-degree, 5) ESA CCI land cover map, 6) Köppen-Kottek classification, 7) Köppen-Peel classification, and 8) FAO ecozone map. (c) Test data R^2 and RMSE grouped into 15-degree latitude zones, using four machine learning models and eight cross-validation methods. The numbers in brackets at the top represent the count of wood density measurements in the test dataset within each latitudinal zone.

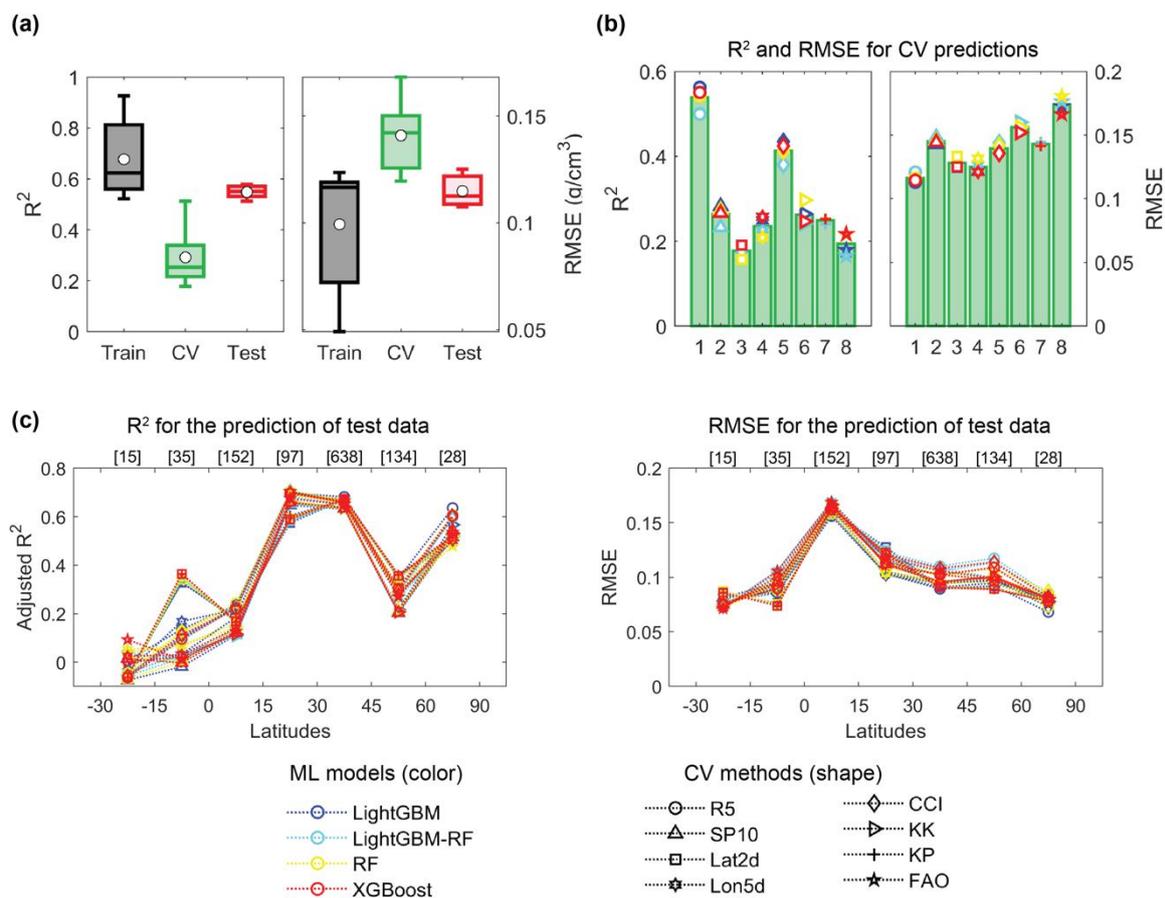
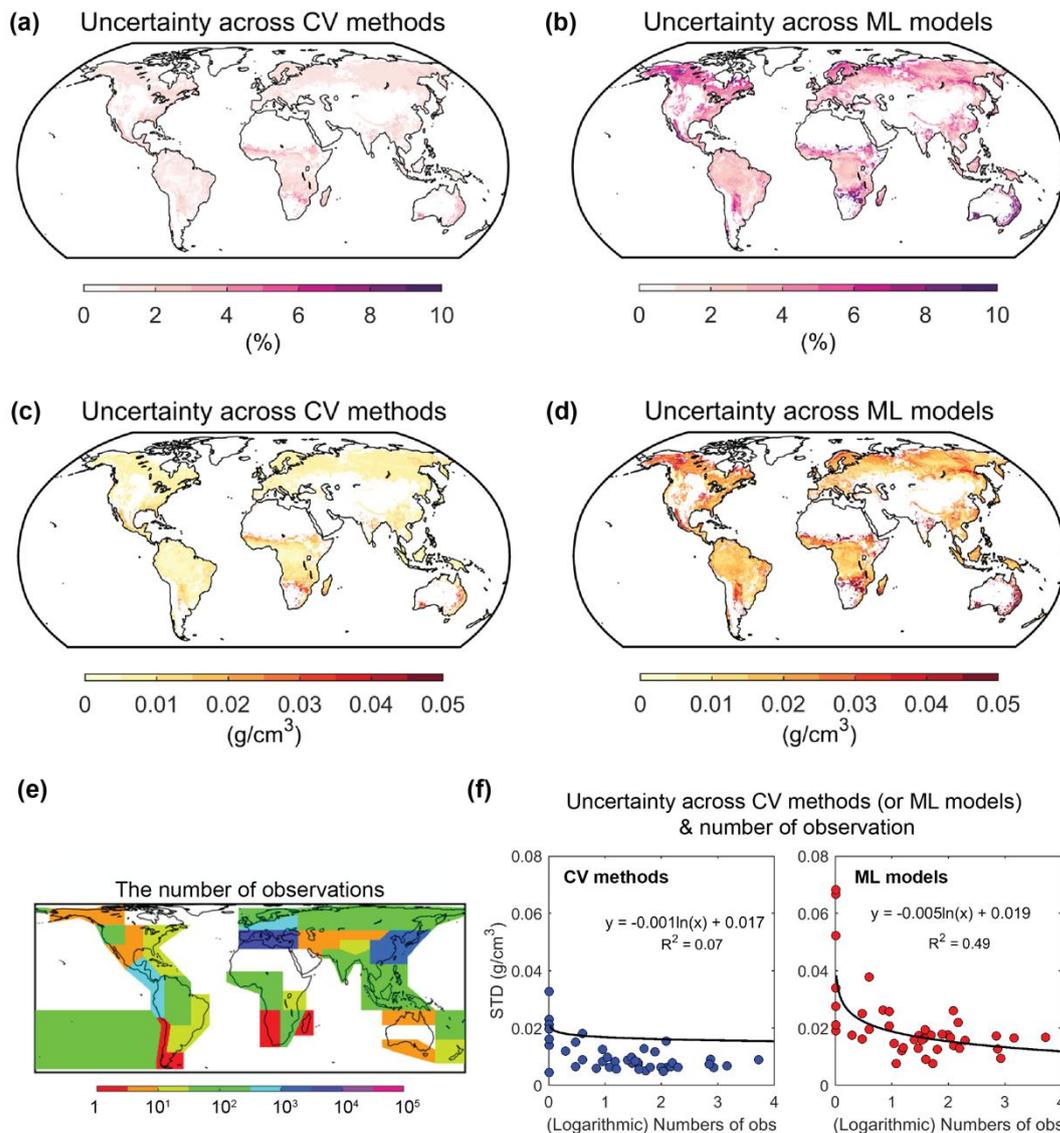


Figure S3. (a)-(d) Spatial patterns of uncertainty in global wood density maps, the uncertainty in (a)-(b) is estimated by the ratio of the standard deviation and mean values, while that in (c)-(d) is the absolute value of standard deviation. The uncertainty is estimated based on two aspects: (a) (c) the standard deviation of wood density generated through eight different cross validation methods, and (b) (d) the standard deviation of estimates derived from four distinct machine learning models. (e) The spatial pattern indicates the quantity of wood density measurement used for the generation of global maps for each IPCC climate reference region. (f) The scatter plots show the relationship between the uncertainty across cross-validation methods (represented by blue dots), as well as machine learning models (represented by red dots) in our predicted wood density estimates and the number of measurements for regions.



(represented by red dots) in our predicted wood density estimates and the number of measurements for regions.

Figure S4. The boxplots show the distribution of wood density for different categories of Köppen climate classification. Both wood density measurement (filled boxes), and our estimates derived from four machine learning models (transparent boxes) are shown. In the plots, the white dot represents the mean value, and the lines outside and inside the boxes represent, from top to bottom, 90th, 75th, 50th, 25th, and 10th percentiles.

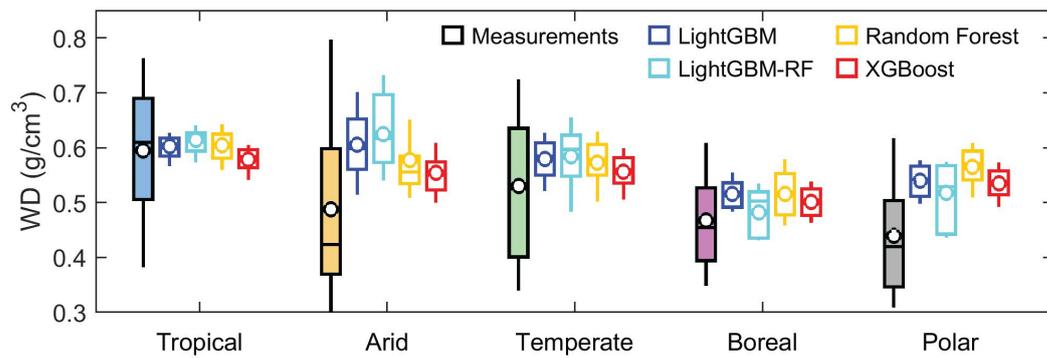


Figure S5. (a) The fraction of NaN values of all the selected features used to predict wood density. (b) The mean of coefficient variation of values from the nearest sites.

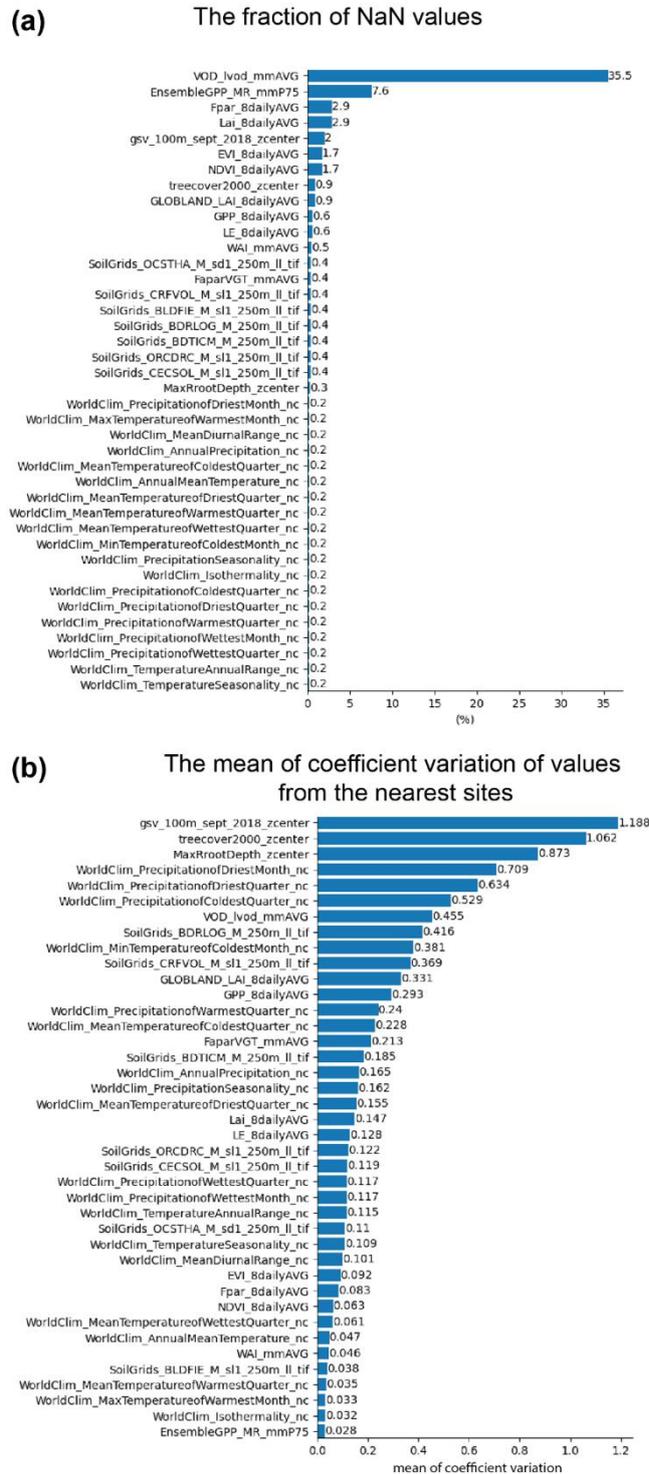


Figure S6. The latitudinal averages (color lines) and standard deviations (color shading) of wood density across a gradient from dry to wet regions. Here, aridity index (AI), calculated as the ratio of precipitation to potential evapotranspiration, was used to illustrate the degree of aridity. $AI < 0.25$: dry ; $0.25 \leq AI < 0.75$: moderate ; $0.75 \leq AI$: wet.

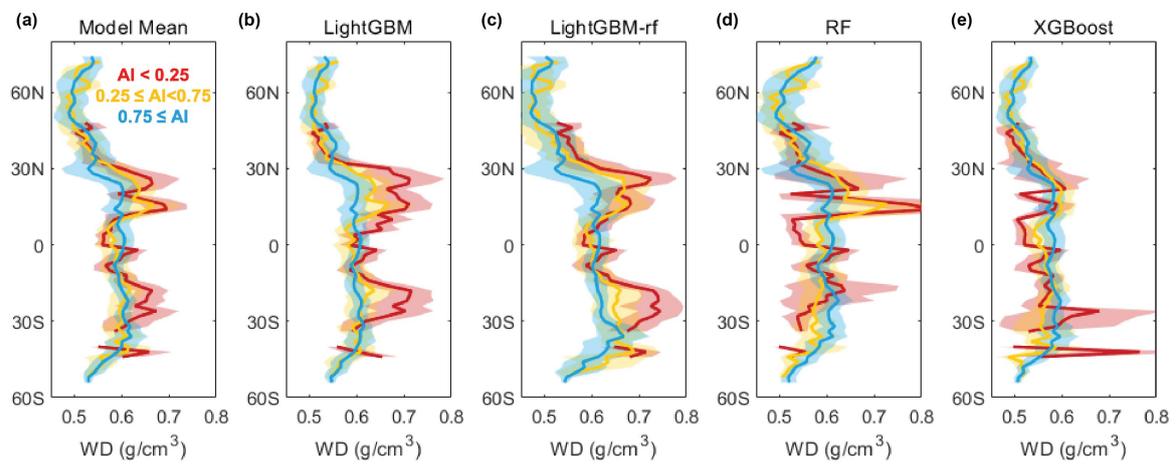


Figure S7. (a) Model performance with (w/) and without (w/o) plant leaf traits: R^2 and root mean squared error (RMSE) for the prediction of training data, cross-validation (CV) data and test data using four distinct machine learning models, i.e., LightGBM, LightGBM-RF, Random Forest and XGBoost and under eight distinct cross-validation methods. (b) Comparison of wood density predictions for test dataset from models with and without plant leaf traits. The dots show the ensemble mean of four machine learning models using eight cross-validation methods are shown. (c) Test data R^2 grouped into 15-degree latitude zones. The solid curve and shading indicate the mean and standard deviation of R^2 from four machine learning models and eight cross-validation methods. The numbers in brackets at the top represent the count of wood density measurements in the test dataset within each latitudinal zone. (d) The feature importance of plant leaf traits in four machine learning models.

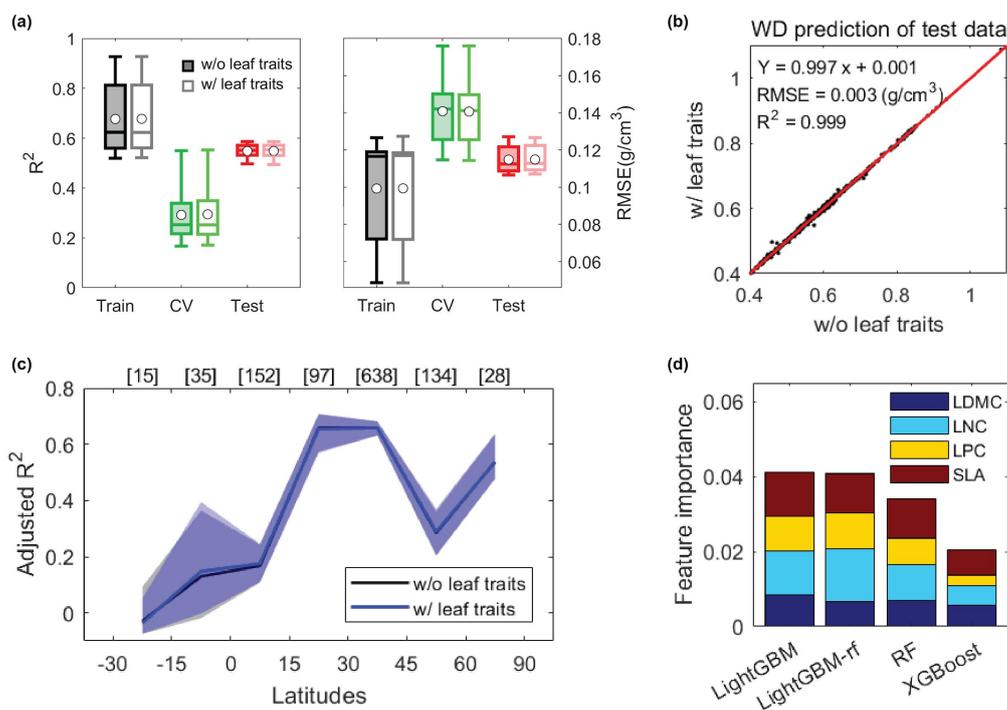


Table S1. A list of databases providing wood density measurements.

No.	Name of database	Reference	Comment
1	ForestPlotNet	Kattge et al. (2020)	from TRY database
2	BAAD 2020		
3	Araucaria forest database		
4	Brazilian database		
5	Bridge database		
6	Catalonian database		
7	Costa Rice dry/traits database		
8	CTFS database		
9	Dinghushan database		
10	FAPESP database		
11	Golfo Dulce database		
12	Jasper Ridge Californian database		
13	LABDENDRO database		
14	Midwestern southern US database		
15	Neotropic Traits database		
16	Netherlands (Plants Traits database		
17	Panama Plant Traits database		
18	Panama wood anatomy database		
19	Pinus Juniperus Traits database		
20	RAINFOR Traits/Plant database		
21	Rehabilitating Coastal database		
22	South Africa wood database		
23	Spanish Traits database		
24	Ukraine Traits/wetlands database		
25	Tropical (Plant) Traits database		
26	Xylem Functional Traits database		
27	Yangambi database		
28	Africa Woody Plants database		
29	Mediterranean Forest Traits database		
30	Brazil Rainforest database		
31	Neotropical Plant Traits database		
32	Functional Traits of Woody Species		
33	Chinese Savanna trees database		
34	Tundra Traits database		
35	New Zealand database		
36	Raja Ampat tree database		
37	Shepaschenko database	Shepaschenko et al. (2017)	Plot-level data
38	Poland wood density database	Personal Communication	Unpublished data
39	UMR AMAP	Personal Communication	Unpublished data

Table S2. The predictor covariates used in the machine learning models for generating global wood density maps. The monthly (8-daily) averages represent where the original monthly (8-daily) values were aggregated into a mean for the entire period.

Variables	Description	Unit	Original resolution	Source
<i>Climate conditions (38)</i>				
MAT	Mean Annual Temperature	°C		
MAP	Mean Annual Precipitation	mm		
TS	Temperature Seasonality, the coefficient variation of monthly temperature (STD/Mean) × 100	%		
PS	Precipitation Seasonality, the coefficient variation of monthly precipitation (STD/Mean)	%		
MTCQ	Mean Temperature of Coldest Quarter	°C		
MTDQ	Mean Temperature of Driest Quarter	°C		
MTWarmQ	Mean Temperature of Warmest Quarter	°C	30 arc sec	WorldClim 2
MTWetQ	Mean Temperature of Wettest Quarter	°C		
MinTCM	Min Temperature of Coldest Month	°C		
MaxTWM	Max Temperature of Warmest Month	°C		
TAR	Temperature Annual Range	°C		
MDR	Mean Diurnal Range (Mean of monthly (max temp – min temp))	°C		
Isothermality	MDR/TAR × 100	%		
MPCQ	Precipitation of Coldest Quarter	mm		

MPDQ	Precipitation of Driest Quarter	mm		
MPWarmQ	Precipitation of Warmest Quarter	mm		
MPWetQ	Precipitation of Wettest Quarter	mm		
MPDM	Precipitation of Driest Month	mm		
MPWM	Precipitation of Wettest Month	mm		
CloudCover_meanannual	Mean annual cloud frequency (%) over 2000-2014	%		
CloudCover_interannualSD	Mean between-year seasonality represented as the mean of the 2000-2014 monthly standard deviations	%	0.0083°	MODCF
CloudCover_intraannualSD	Within-year seasonality represented as the standard deviation of mean 2000-2014 monthly cloud frequencies	%		
LWdown	Downward Longwave Radiation, daily mean	W/m ²		
LWdown_annP75	Downward Longwave Radiation, Q ₇₅ of annual values	W/m ²	0.5°	CERES
LWup	Upward Longwave Radiation, daily mean	W/m ²		
SWdown	Downward Shortwave Radiation, daily mean	W/m ²		
SWup	Downward Shortwave Radiation, Q ₇₅ of annual values	W/m ²	0.5°	CERES
SWup_std	Upward Shortwave Radiation, daily mean	W/m ²		
Rn	Net Radiation, daily mean	W/m ²		
VPDday	Daily vapour pressure deficit, monthly mean	Pa		
VPDday_mmSTD	Daily vapour pressure deficit, standard deviation of monthly values	Pa	0.5°	ERA interim

VPDday_mmP05	Daily vapour pressure deficit, Q ₀₅ of monthly values	Pa		
VPDday_mmP25	Daily vapour pressure deficit, Q ₂₅ of monthly values	Pa		
VPDday_mmP75	Daily vapour pressure deficit, Q ₇₅ of monthly values	Pa		
WAI	Water Availability Index	mm		
PET_PT	Potential evapotranspiration, the Priestley-Taylor (PT) equation, daily average	mm		
PET_PT_annP05	Potential evapotranspiration, the Priestley-Taylor (PT) equation, Q ₀₅ of annual values	mm		
Rainy days	Rainy days	days	0.05°	Climate Hazards Group InfraRed Precipitation with Station data(CHRIPS)
<i>Soil properties (20)</i>				
BDRLOG	Probability of occurrence of R horizon	%		
BDTICM	Absolute depth to bedrock	cm		
BLDFIE	Bulk density	dg/m ³		
CECSOL	Cation Exchange Capacity of soil	cmolc/kg		
CLYPPT	Weight percentage of the clay particles (<0.0002 mm)	%	250 m	SoilGrids database
CRFVOL	Volumetric percentage of coarse fragments (>2 mm)	%		
OCSTHA	Soil organic carbon stock	ton/ha		
ORCDRC	Soil organic carbon content	permille		
PHIHOX	pH index measured in water solution	pH		

PHIKCL	pH index measured in KCI solution	pH		
SLTPPT	Weight percentage of the silt particles (0.0002–0.05 mm)	%		
SNDPPT	Weight percentage of the sand particles (0.05–2 mm)	%		
AWCh1	Available soil water capacity (volumetric fraction) with FC = pF 2.0	%		
AWCh2	Available soil water capacity (volumetric fraction) with FC = pF 2.3	%		
AWCh3	Available soil water capacity (volumetric fraction) with FC = pF 2.5	%		
WWP	Available soil water capacity (volumetric fraction) until wilting point	%		
AWCtS	Saturated water content (volumetric fraction) for total soil	%		
SM	Soil moisture from 2011 to 2017, monthly averages	m ³ /m ³	0.25°	SMOS IC
<i>Vegetation properties (15)</i>				
PALSAR_HH_zcenter	ALOS Phased Array type L-band Synthetic Aperture Radar (PALSAR) Polarization data, mean value	-		
PALSAR_HH_zstd3×3	ALOS PALSAR Polarization data, standard deviation of polarization data from 3×3 spatial window		10m	JAXA
GSV	growing stock volume for the year 2010	m ³ /ha	100m	GlobBiomass
Tree Cover	Tree cover in the year 2000, defined as canopy closure for all vegetation taller than 5m in height.	%	30m	Hansen Global Forest Watch

EVI	8-daily Enhanced Vegetation Index (EVI) generated using the gridded daily surface reflectance product.	1	0.083°	MOD13A2
NDVI	8-daily Normalized Difference Vegetation Index (NDVI) generated using the gridded daily surface reflectance product.	1		
GPP	Gross Primary Production from 2001 to 2015, 8 daily averages	g/m ² /day	0.01°	High resolution FLUXCOM Ensemble
LE	Latent heat from 2001 to 2015, 8 daily averages	kJ/kg		
Fpar	Fraction of Photosynthetically Active Radiation (FPAR)	%	0.083°	MOD15A2
LAI	Leaf Area Index (LAI)	m ² /m ²		
Globland LAI	Leaf Area Index (LAI) from 2001 to 2015, 8 daily averages	m ² /m ²	0.072°	GlobMAP product
Ensemble GPP mmP75	Gross Primary Production from 1982 to 2011, the 75 th percentile of monthly values	g/m ² /day	0.5°	Middle resolution FLUXCOM GPP Ensemble
FaparVGT	Radiation absorbed by the vegetation (FAPAR) from 1999-2012, monthly averages	%	0.05°	Fapar VGT BioPar product
LVOD	Vegetation Optical Thickness at Nadir from 2011 to 2017, monthly averages	1	0.25°	SMOS IC
Rooting depth	Maximum root depth	m	0.0083°	Fan et al. (2017)
<i>Categorical variables (6)</i>				
Leaf Type	Two levels. B: Broadleaves, N: Needleleaves			
Leaf Habit Type	Two levels. E: Evergreen, D: Deciduous		300m	CCI land cover
Leaf type & Leaf habit type	Four levels. EBF, ENF, DBF, DNF			
FAO ecozone	Ecological zones (20 classes)		0.083°	FAO
Koepfen Geiger	Climate classification (31 classes)		0.5°	Kottek et al. (2006)
MODIS LC	Land cover types (IGBP classification)		0.012°	MCD12Q1