A global map of root biomass across the world's forests

Yuanyuan Huang[1,2*], Phillipe Ciais[1], Maurizio Santoro[3], David Makowski[4,5], Jerome Chave[6], Dmitry Schepaschenko[7,8,9], Rose Abramoff[1], Daniel S. Goll[10], Hui Yang[1], Ye Chen[11], Wei Wei[12], Shilong Piao[13,14,15]

[1]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France.

[2]Commonwealth Scientific and Industrial Research Organisation, Aspendale, 3195, Victoria, Australia.

[3]Gamma Remote Sensing, 3073 Gümligen, Switzerland.

[4]INRA, AgroParisTech, University Paris-Saclay, UMR 211, F-78850 Thiverval-Grignon, France.

[5]CIRED, 45bis Avenue de la Belle Gabrielle, 94130 Nogent-sur-Marne, France.

[6]Laboratoire Evolution et Diversite Biologique UMR 5174, CNRS, Universite Paul Sabatier, 118 route de Narbonne, Toulouse, 31062 France.

[7]International Institute for Applied Systems Analysis (IIASA) Schlossplatz 1, A-2361 Laxenburg, Austria.

[8]Center of Forest Ecology and Productivity of the Russian Academy of Sciences, Moscow 117997, Russia

[9]Siberian Federal University, Krasnoyarsk, 660041, Russia

[10]Department of Geography, University of Augsburg, Germany.

[11]Department of Mathematics and Statistics, Northern Arizona University, 86001, Flagstaff, AZ, US.

[12]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-environmental Sciences, Chinese Academy of Sciences, Beijing, 100085, China.

[13]Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing, China

[14]Key Laboratory of Alpine Ecology and Biodiversity, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

[15]Center for Excellence in Tibetan Earth Science, Chinese Academy of Sciences, Beijing, China

32 *Corresponding author: Yuanyuan Huang

33 Email: Yuanyuan.Huang@csiro.au ; Phone: + 61392394406

Abstract (150 words limits)

Root plays a key role in plant growth and functioning. Here we combine 10307 field measurements of forest root biomass worldwide with global observations of forest structure, climatic conditions, topography, land management and soil characteristics to derive a spatially-explicit global high-resolution (~ 1km) root biomass dataset, including fine and coarse roots. In total, 142 ± 32 Pg of live dry matter biomass is stored below-ground, that is a global average root:shoot biomass ratio of 0.25 ± 0.10. Our estimations of total root biomass in tropical, temperate and boreal forests are 44-226% smaller than earlier studies[1-3]. The smaller estimation is attributable to the updated forest area, spatially explicit above-ground biomass density used to predict the patterns of root biomass, new root measurements and upscaling methodology. We show specifically that the root shoot allometry is one underlying driver that leads to methodological overestimation of root biomass in previous estimations.

Keywords

Root; Carbon; Machine Learning; Forest; Biomass; Globe

Roots mediate nutrient and water uptake by plants, below-ground organic carbon decomposition, the flow of carbohydrates to mycorrhizae, species competition, soil stabilization and plant resistance to windfall[4]. The global distribution of root biomass is related to how much photosynthates plants must invest below-ground to obtain water, nitrogen and phosphorus for sustaining photosynthesis, leaf area and growth. Root biomass and activity also control the land surface energy budget through plant transpiration[4,5]. While Earth Observation data combined with field data enables the derivation of spatially explicit estimates of above-ground biomass with a spatial resolution of up to 30 meters over the whole globe[6,7], the global carbon stock and spatial details of the distribution of below-ground root biomass (fine + coarse) relied on punctual measurements and coarse extrapolation so far, therefore remaining highly uncertain

More than twenty years ago, Jackson et al, 1996, 1997 [1,2] provided estimates of the average biomass density (weight per unit area) and vertical distribution of roots for 10 terrestrial biomes. Multiplying their average root biomass density with the area of each biome gives a global root biomass pool of 292 Pg, with forests accounting for ~68% of it. Saugier, et al. (2001) estimated global root biomass to be 320 Pg by multiplying biome-average root to shoot ratios (*R:S*) by shoot biomass density and the land area of each biome. Mokany, et al. (2006) argued that the use of mean *R:S* values at biome scale is a source of error because root biomass measurements are performed at small scales with roots having a high spatial heterogeneity and their size distribution spanning across several orders of magnitude, the fine roots being particularly difficult to sample[8,9]. With updated *R:S* and broader vegetation classes, they gave a higher global root biomass of 482 Pg. Robinson (2007) further suggested a 60% underestimation of *R:S*, which translated into an even higher global root biomass of 540-560 Pg. These studies provided a first order estimation of the root biomass for different biomes, but not of its spatial details and it is worth noting that numbers have increased with time.

An alternative approach to estimate root biomass is through allometric scaling, dating back to West, Brown and Enquist (1997, 1999)[6 7] and Enquist and Niklas (2002). The allometric scaling theory assumes that biological attributes scale with body mass, and in the case of roots, an allometric equation verified by data takes the form of $R \propto S^{\beta}$ where R is the root mass, S the shoot mass and β a scaling exponent. Differently than in the studies listed above assuming the R:S ratio to be uniform, this equation implies that the *R:S* ratio varies with shoot size as β is not equal to one [10-15] . Allometric equations also predict that smaller trees generally have a larger

R:S with $\beta < 1$, which is well verified by measurement of trees of different sizes [12-15]. The allometric equation approach was applied for various forest types, and the scaling exponent β was observed to differ across sites[16], species[17], age[13], leaf characteristics[18], elevation[19], management status[20], climatic conditions, such as temperature[21], soil moisture and climatic water deficit [20], as well as soil nutrient content and texture[14]. Despite successful application of allometric equations for site- and species-specific studies[16], their use to predict large-scale and global root biomass patterns appears to be challenging.

Here we use a new approach to upscale root biomass of trees at global scale based on machine learning algorithms trained by a large dataset of measurements and using as predictors high-resolution maps of tree density, above-ground biomass, soils and environmental drivers (Supplementary Tables 1, 2). Firstly, we collected 10307 *in-situ* measurements [14,30,31]of the biomass of roots and shoots for individual woody plants (see Methods, Supplementary Data), covering 465 species across 10 biomes defined by The Nature Conservancy[22] (Supplementary Figure 2). In biomes like savannas where trees and woody plants can be sparse, we estimate root biomass as the average for the woody plants present in that biome given a canopy cover threshold of 15% at a 30 m resolution globally[23]. In the root below-ground biomass estimates (BGB) we count both coarse and fine roots. We acknowledge the importance of understanding large scale temporal dynamics of fine root. As a first step, this study aims at the spatial pattern of total root biomass. We upscaled root biomass from individual plant level measurements rather than from stand-level data because a large number of primary data are collected for individual woody plants, and this approach allows us to account in both the training of machine learning models and their upscaling results, for the fact that root biomass depends on tree size or above-ground biomass[14,15,20]. We searched through a pool of 47 predicting variables that include above-ground biomass and other vegetation variables, edaphic, topographic, anthropogenic and climatic conditions (Supplementary Table 1). Different machine learning models were tested, and we selected the model that performs best on cross validation samples (see Methods for model selection criterion). The best model is a random forest (RF, see Methods) and we mapped global root biomass at a 1 km resolution through this model relying on 14 predicting gridded variables, including the shoot biomass of an average tree derived from shoot density (weight per area) [24] and tree density (number of trees per area)[25], tree height[26], soil nitrogen[27], pH[27], bulk density[27], clay content[27], sand content[27], base saturation[27], cation exchange capacity[27], water vapor

pressure[28], mean annual precipitation[28], mean annual temperature[28], aridity[29] and water table depth[30] (see Supplementary Table 1 for detailed information and references). To estimate root biomass pools at global and biome scales, the mean root biomass of trees in each 1 km pixel was multiplied by a tree density map available at the resolution of 1 km from ref. [25] (see Methods)

Results

We estimated a global total root biomass of 142±32 Pg (see Method for uncertainty estimation and Supplementary Figures 3, 4) when forest is defined as all areas with tree cover larger than 15% from the Hansen et al. (2013) tree cover map. The corresponding global weighted mean *R:S* is 0.25 ± 0.10. The root biomass spatial distribution generally follows the pattern of shoot biomass, but there are significant local and regional deviations as shown by Figure 1. 51% of the global tree root biomass comes from tropical moist forest, 14% from boreal forest, 12% from temperate broadleaf forest and 10% from woody plants in tropical and subtropical grasslands, savanna and shrublands (Supplementary Table 3). Given our use of a tree cover threshold of 15% at 30m resolution, our estimate ignores the roots of isolated woody plants present in arid or cold regions [31], as well as heterogeneous (e.g. urban or agriculture) landscapes and is possibly an under-estimate. Total root biomass decreases from 151 to 134 Pg when the canopy cover threshold used to define forest land is increased from 0% to 30%. The root biomass density per unit of forest area is highest in tropical moist forest, followed by temperate coniferous and Mediterranean forest (Figure 1, Supplementary Table 3). Cross validation showed a good match between predictions from our RF model and *in-situ* observations (Figures 2e, all data; Supplementary Figure 6, for each biome), with a coefficient of determination R^2 of 0.85 and a median *R:S* similar to validation samples (0.35 from *in-situ* observation vs. 0.38 from prediction). Root biomass of tropical, temperate and boreal forests together is 44-226% lower compared to earlier studies (Table 1, Supplementary Table 5, see Supplementary Information Comparison with published results).

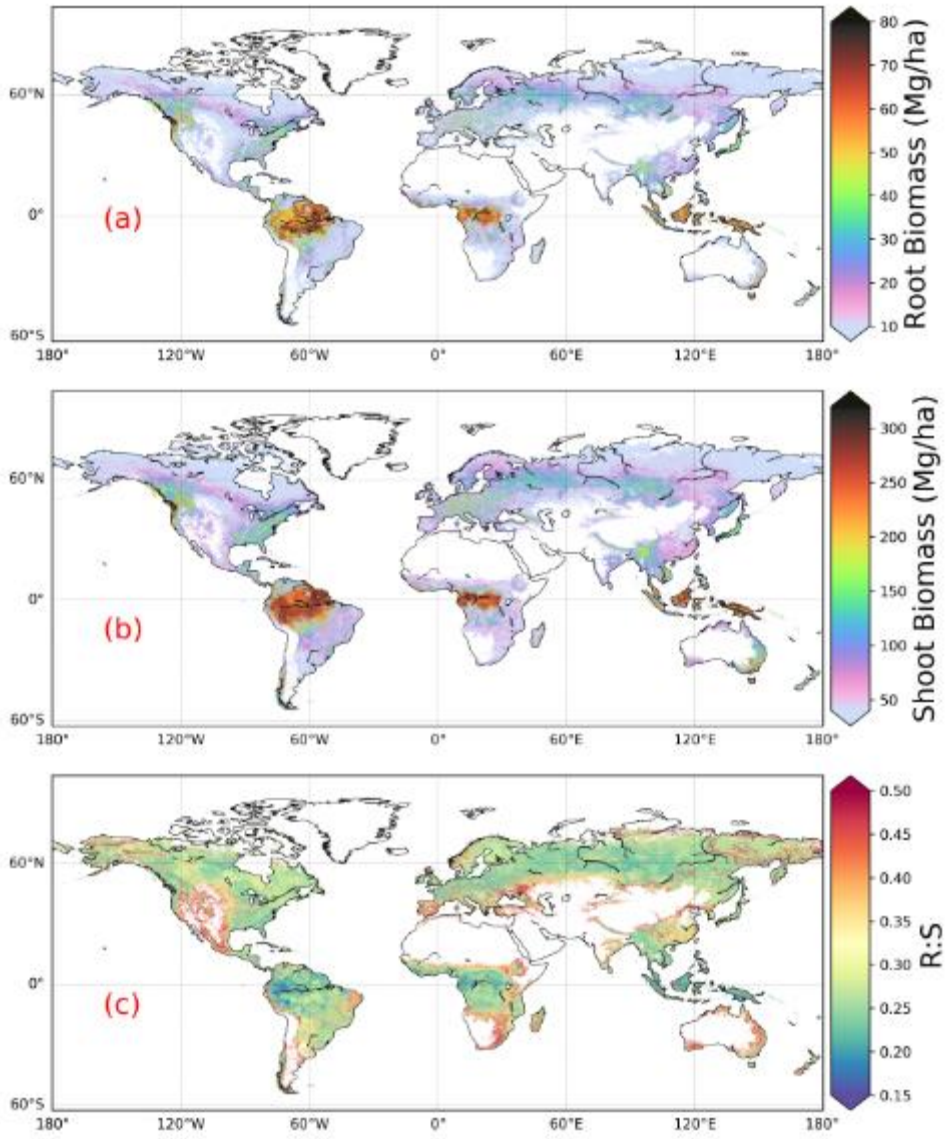

Figure 1. Global maps of forest root biomass generated through the random forest model (a), shoot biomass from GlobBiomass-AGB[6] (b) and *R:S* (c). Forest is defined as an area with canopy cover > 15% from the Hansen et al. (2013) tree cover map.

We then analysed the dominant factors explaining spatial variations of root biomass and *R:S* (see Methods). Broadly speaking, locations with small trees, low precipitation, strong aridity, deep water table depth, high acidity, low bulk density, low base saturation and low cation exchange capacity are more likely to have higher fractional root biomass (Figure 3). In line with the allometric theory, shoot biomass emerged as the most important predictor of *R:S* and root

biomass, as given by the Spearman correlation analysis shown in Figure 3, and partial importance plots (Supplementary Figures 7, 8, 9). Water related variables (precipitation, water table depth, aridity and vapor pressure) also emerged as important predictors in explaining *R:S* patterns (Figure 3)[20], with trees and woody plants in dry regions generally having higher *R:S* (Supplementary Tables S3, S4), and with stronger dependence on precipitation when it is small and on water table depth when it is deep. Temperature is slightly negatively correlated with *R:S* at the global scale, in line with Reich et al. (2014). However, the relationship between temperature and below-ground biomass is not consistent among biomes (Figure 3) and biomass size groups (Supplementary Figures 7, 8, 9). The relationship between total soil nitrogen and root biomass is negative when soil nitrogen content is below 0.1-0.2 % (Supplementary Figure 7, 8, 9). Root biomass and *R:S* generally increases with soil alkalinity (Figure 3, Supplementary Figures 7, 8, 9). Low pH is toxic to biological activities and roots, especially fine roots are sensitive to soil acidification, as revealed by a recent meta-analysis[32]. Our results also indicate overall positive correlations between CEC, BS and R:S, but the processes that may account for these correlations are less clear from literature. Age has been shown to be important for R:S[33]. How age regulates *R:S* remains elusive, with studies showing positive[34], slightly negative[35] or no relationship[36] between *R:S* and age. Including forest age (see Methods: Preparing predicting variables) as a predictor only marginally improved our model prediction (see SI for details). It is likely that shoot biomass partially accounts for age information and the quality of the global forest age data might also affect the power of this variable in improving root biomass predictions.

Table 1. Comparison between studies quantifying root biomass in tropical, temperate and boreal forests.

	This study[S1]	This study[S2]	Jackson[1]	Saugier[2]	Robinson[3]	This study [S3]
Method	Machine learning	Machine learning	Biome average root biomass density	Biome average *R:S*, shoot biomass density	Biome average *R:S*, shoot biomass density	Allometric equations
Tropical (Tr, Pg)	92	76	114	147	246	
Temperate(Te, Pg)	26	25	51	59	98	
Boreal (Bo, Pg)	21	20	35	30	50	
Tr + Te + Bo (Pg)	139	121	200	236	394	
Globe (Pg)	142	142				155-210
RD_{S1}[*]	0%		44%	70%	183%	
RD_{S2}[&]		0%	65%	95%	226%	

S1, Tropical moist forest (Biome 1), tropical dry forest (Biome 6), tropical/subtropical coniferous forest (Biome 11) and forest in tropical/subtropical grasslands/savannas and shrublands (Biome 3) are aggregated to represent tropical systems (Tr). Temperate broadleaf/mixed forest (Biome 4), temperate coniferous forest (Biome 5) and forest in

temperate grasslands/savannas and shrublands (Biome 8) are merged together as temperate systems (Te). Boreal forest (Biome 2) and woody plants in tundra region (Biome 7) are aggregated as boreal forest (Bo). Biome classification is from The Nature Conservancy[22] and is shown in Supplementary Figure 2.

S2, Tropical systems (Tr): Biomes 1,6,11; Temperate systems (Te): Biomes 4,5; Boreal systems (Bo): Biome 2.

S3, Estimation based on allometric equations and the global above-ground biomass dataset from ref[6]. See Supplementary Table 7 for details.

* RD_{S1}, the relative difference of Tr + Te + Bo between this study (S1) and previous quantifications. RD_{S1} = (previous study – this study)/this study x 100%. For example, in the column with the head Jackson, RD_{S1} = (200-139)/139*100% = 44%.

& RD_{S2}, the same as RD_{S1}, but with the S2 definition of tropical, temperate and boreal systems.

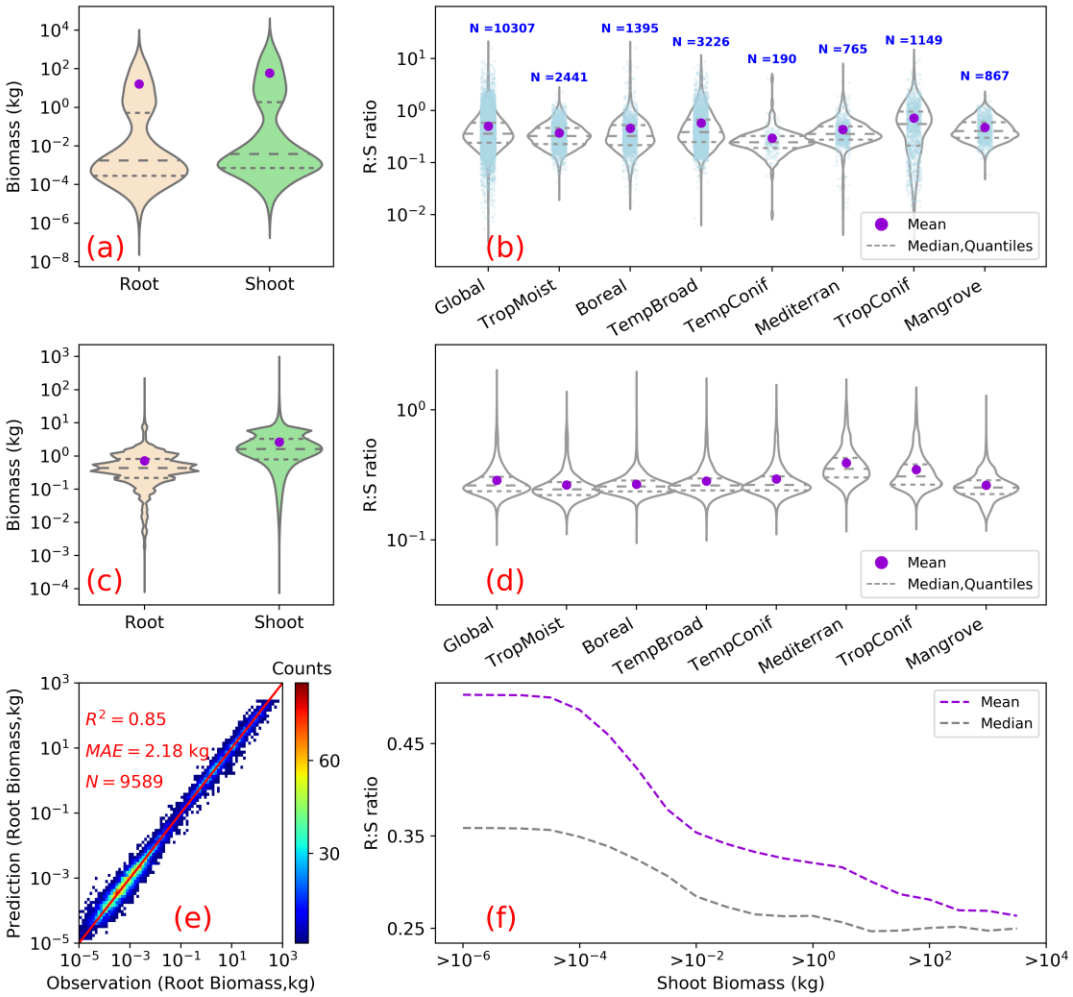

Figure 2. Root biomass and root shoot ratio (*R:S*). (a) and (b) show as violin plots the distribution of root and shoot biomass (in unit of kg/plant) and *R:S* ratios in the raw data used for upscaling. (c) and (d) are the distributions of model predicted root biomass from this study, of above-ground biomass used for the predicting, and of modelled *R:S* ratios at the global and biome scales. (e) is a heat plot of observed vs. predicted root biomass in kg of root per individual woody plant. (f) shows the mean (purple) and median (grey) *R:S* as a function of shoot biomass from observations. A shift of the shoot biomass towards a larger size ((a), (c)) results in a smaller predicted mean *R:S* at the global scale ((b),(d)) (see Supplementary Table 4 for exact values) as

the mean *R:S* is size dependent (f). R^2 is the coefficient of determination, *MAE* is the mean absolute error and *N* is the number of samples. TropMoist: tropical moist forest; Boreal: boreal forest/taiga; TempBroad: temperate broadleaf and mixed forest; TempConif: temperate coniferous forest; Mediterran: Mediterranean forests, woodlands and scrub; TropConif: tropical and subtropical coniferous forest; and Mangrove forest: mangrove forest. Note that the scales of y-axis are different between (a) and (c), (b) and (d). Model training and prediction were conducted on filtered data with *R:S* falling between the 1st and 99th percentiles and shoot biomass matching the range derived from GlobBiomass-AGB[6] to reduce impacts from outliers.

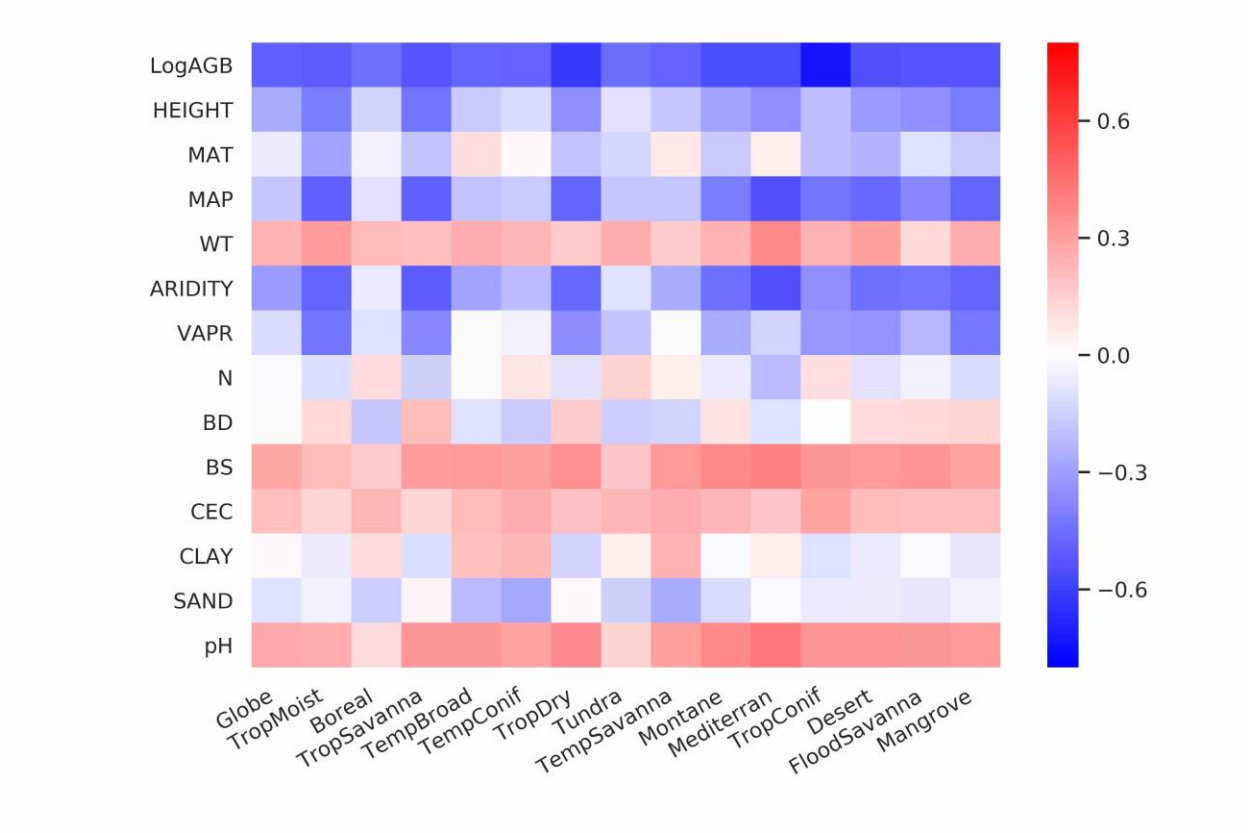

Figure 3. Spearman rank correlations between predicting variables and log-transferred *R:S*. Spearman coefficients are shown at both the global and biome scales for LogAGB: the logarithm of shoot biomass with base 10; HEIGHT: plant height; MAT: mean annual temperature; MAP: mean annual precipitation; WT: water table depth; ARIDITY: the aridity index; VAPR: water vapor pressure; N: soil nitrogen content; BD: soil bulk density; BS: soil base saturation; CEC, soil cation exchange capacity; CLAY: soil clay content; SAND: soil sand content; and pH: soil pH. From left to right, biomes are ordered descendingly according to their forest areas (Supplementary Figure 2).

Discussion

Our lower estimation of root biomass compared to earlier studies is attributable to differences in forest area (Supplementary Table 5), above-ground biomass density (Supplementary Table 5), root biomass measurement and upscaling methodology. For example, the forest area in temperate zones used in Jackson et al. (1997) was about one third higher than in this study, which partly explains their higher root biomass for this biome (Supplementary Table 5). Our lower values of root biomass compared to Saugier et al. (2001), Mokany et al. (2006) and Robinson (2007) are caused mainly by our lower above-ground biomass density and *R:S* (Supplementary Table 5). Shoot biomass density (AGB) of tropical zones is 70% lower in our study than in Robinson (2007) who used sparse plot data collected more than a decade ago (Supplementary Table 5, case S2), and this lower AGB explains 27-46% of our lower root biomass (Supplementary Tables 5, 6). On the other hand, lower biome average *R:S* explains 41-48% of our underestimation compared to Robinson (2007). To elucidate this difference, we calculated weighted biome average *R:S* ratios through dividing total biome level shoot biomass by root biomass (i.e., weighted mean *R:S*). These weighted mean *R:S* ranging between 0.19 and 0.31 across biomes (Supplementary Table 3) are generally smaller than the *R:S* values reported in previous studies, which were based on average ratios obtained from sparser data (Supplementary Table 5), despite the arithmetic mean *R:S* (without weighting by biomass) from woody plants located in tropical, temperate and boreal zones (Supplementary Table 4) from our database being close to those from Robinson (2007).

The common practice of estimating root biomass through an average *R:S* without considering the high spatial variability of biomass and this ratio[4] is a source of systematic error, leading to overestimating the global root biomass for two reasons. Firstly, upscaling ratios through arithmetic averages (possibly weighted by the number of trees or area, but not accounting for the fine grained distribution of biomass) systematically overestimates the true mean *R:S* (see SI Arithmetic mean *R:S* section) because *R:S* is a convex negative function of S given by $R{:}S \propto S^{\beta-1}$ with β taking typical values of about 0.9 [35,37,38]. This explains why high-resolution S data used to diagnose weighted mean *R:S* ratios in our approach give generally smaller values than using arithmetic means at the biome level (see also Supplementary Tables 3 and 4). Secondly, available measurements tend to sample more small woody plants than big trees compared to real world distributions, because small plants are easier to excavate for measuring

roots (see Figure 2a, 2c) but smaller plants tend to have larger *R:S* (Figure 2e). This sampling bias shifts the *R:S* towards larger values when using the mean from all samples in current databases. Our RF approach uses these data for training but in the upscaling, it accounts for realistic distributions of plant size. We further verified that our upscaled *R:S* ratios are robust to sub-sampling the training data in observed distributions, so that the bias of training data towards small plants does not translate into a bias of upscaled results.

The upscaling approach using allometric equations should also tend to overestimate the global root biomass due to the curvature of these allometric functions (see SI Allometric upscaling section). The global forest root biomass ranges between 154 – 210 Pg when root biomass was upscaled through different allometric equations collected from literature and fitted to our database (Supplementary Table 7), generally larger than from the RF mapping. Excluding the under-sampling issue in root biomass measurement, the global root biomass is likely to be smaller than when applying the allometric equation to the spatial average of shoot biomass (Supplementary Figures 10,11,12,13). Thus, future *in-situ* characterization of size structure across the world's forests (see SI Allometric upscaling section) would greatly improve root biomass quantification.

An accurate spatially explicit global map of root biomass helps to improve our understanding of the Earth system dynamics by facilitating fundamental studies on resource allocation, carbon storage, plant water uptake, nutrient acquisition and other aspects of biogeochemical cycles. For example, the close correlation (correlation coefficient: 0.8) between root biomass and rooting depth[39] at the global scale and the importance of root in plant water uptake and transpiration reflect close interactions between vegetation and hydrological cycles. The quest for drivers that affect allocation and consumption of photosynthetic production is a major focus of comparative plant ecology and evolution, as well as the basis of plant life history, ecological dynamics and global changes[11]. Turnover time and allocation are two key aspects that contribute to large uncertainties in current terrestrial biosphere model predictions[40,41]. Our root biomass map does not provide data on turnover or allocation, but an outcome on their aggregated effects. Future studies combining the root biomass map with upscaled root turnover data could shed light on the allocation puzzle. The growth of the fast turnover part of root, mostly fine root, and leaf are highly linked. If we assume an annual turnover of leaf and fine root, a preliminary estimation of average forest fine root biomass (from leaf biomass) reaches 6.7-7.7 Pg (see

Supplementary Information: Preliminary estimation of fine root biomass). Despite being a small portion and highly uncertain, fine roots are temporally variable and functionally critical in ecosystem dynamics. Future studies on global distribution and temporal dynamics of fine roots are valuable. Considering specific biomes, tropical savannas would benefit from better root biomass estimation due to its large land area, and in tropical dry forests, field measurements of root and shoot biomass are needed to refine root biomass quantifications.

Methods

Overview

Our global mapping of root biomass relies on a predicting model based on a machine learning algorithm that is fitted to a large number of ground field measurements. Root biomass was upscaled as a function of shoot biomass, tree height, age, species, land management, topography, edaphic and climate variables. The process takes three major steps (Supplementary Figure 1). The first step is to collect field measurements, and observations of auxiliary variables such as tree height, age, species and management status (see sections field measurements and preparing predicting variables below). In a second step, we compared the allometric upscaling and tested three machine learning techniques, the random forest (RF), the artificial neural networks (ANN) and multiple adaptive regression splines (MARS) through 47 input variables. The best predicting model with the minimum number of predictors and with the lowest mean absolute error (MAE) and highest R-squared value (R^2) was selected through cross-validation (see section Building predicting models below). The next step was to generate a 1 km global root biomass map by running the best predicting model on spatially-explicit gridded fields of model inputs. The model outputs were initially expressed as root biomass in unit of weight per individual woody plant and were then mapped into root biomass per unit area using tree densities (the number of trees per unit area)[25]. The uncertainty of the mapping and the importance of the model inputs were analysed in detail as explained below.

Field measurements

Our dataset was compiled from literature and existing forest biomass structure or allometry databases[42 20,33,43]. We included studies and databases that reported georeferenced location, root biomass and shoot biomass. For example, Ref[44] is not included due to lack of

georeferenced location and Ref[45] in not used as we also need measurements of other plant compartments like shoot biomass. Repeated entries from existing databases were removed. One of the databases[42] reported data on woody plants which also include shrub species. We kept the shrub data partly because the remote sensing products we used to generate our root map do not clearly separate trees from shrubs. Around 82% of the extracted entries also recorded plant height and management status. Height was identified as an important predictor in our model assessment, and entries were discarded when height was missing (18% of data). As woody plant age was reported in 19% of the entries only, the values of this variable was determined from another source of information, i.e. from a composite global map introduced in the next section. Species names were systematically reported, but biotic, climatic, topographic and soil information were missing for a substantial proportion of entries and values of these variables were thus extracted from independent observation-driven global maps as explained in the next section. Our final dataset includes biomass measurements collected in 494 different locations from 10307 individual plants, which cover 465 species across 10 biomes as defined by The Nature Conservancy[22] (Supplementary Figure 2; Supplementary Data).

Preparing predicting variables

We used 47 predictors that broadly cover 5 categories: vegetative, edaphic, climatic, topographic and anthropogenic (Supplementary Table 1). Vegetative variables include shoot biomass, height, age, maximum rooting depth, biome class and species. Edaphic predictors cover soil bulk density, organic carbon, pH, sand content, clay content, total nitrogen, total phosphorus, Bray phosphorus, total potassium, exchangeable aluminium, cation exchange capacity, base saturation (BS), soil moisture and water table depth (WT). Climatic predictors are mean annual temperature (MAT), mean annual precipitation (MAP), the aridity index that represents the ratio between precipitation the reference evapotranspiration, solar radiation, potential evapotranspiration (PET), vapor pressure, cumulative water deficit (CWD=PET - MAP), wind speed, and mean diurnal range of temperature (BIO2), isothermality (BIO2/BIO7) (BIO3), temperature seasonality (BIO4), max temperature of warmest month (BIO5), min temperature of coldest month (BIO6), temperature annual range (BIO7), mean temperature of wettest quarter (BIO8), mean temperature of driest quarter (BIO9), mean temperature of warmest quarter (BIO10), mean temperature of coldest quarter (BIO11), precipitation of wettest month (BIO13),

346 precipitation of driest month (BIO14), precipitation seasonality (BIO15), precipitation of wettest
347 quarter (BIO16), precipitation of driest quarter (BIO17), precipitation of warmest quarter
348 (BIO18), precipitation of coldest quarter (BIO19). The topographic variable is elevation and we
349 take the management status (managed or not) as the anthropogenic predictor. All references are
350 given in Supplementary Table 1.
351 To derive the shoot or above-ground biomass (AGB) per tree (in unit of weight per tree),
352 we combined the GlobBiomass-AGB satellite data product[24] (in unit of weight per unit area)
353 with a tree density map (number of trees per unit area)[25]. The GlobBiomass dataset was based on
354 multiple remote sensing products (radar, optical, LiDAR) and a large pool of *in-situ* observations
355 of forest variables[6,46]. The original GlobBiomass-AGB map was generated at 100 m spatial
356 resolution; for this study, the map was averaged into a 1 km pixel by considering only those
357 pixels that were labeled as forest [6]. A pixel was labeled as forest when the canopy density was
358 larger than 15% according to Hansen et al. (2013)'s dataset (Hansen2013) averaged at 100 m.
359 The 1-km resolution global tree density map was constructed through upscaling 429,775 ground-
360 based tree density measurements with a predictive regression model for forests in each biome[25].
361 The forest canopy height map took advantage of the Geoscience Laser Altimeter System (GLAS)
362 aboard ICESat (Ice, Cloud, and land Elevation Satellite). Forest definitions are slightly different
363 among these three maps. Forest area of the tree density map was based on a global consensus
364 land cover dataset that merges four land cover products [47], which gave an equal total tree count
365 as the Hansen et al. (2013) land cover [25]. The canopy height map used the Globcover land cover
366 map[48] as reference to define forest land. We took Hansen2013 with a 15% canopy cover
367 threshold as our base forest cover map. We approximated the missing values in tree density and
368 height (due to mismatches in forest cover) by the mean of a 5x5 window that is centered on the
369 corresponding pixel. We quantified the potential impact of mismatches in forest definition by
370 looking into two different thresholds: 0% and 30%.
371 We merged several regional age maps to generate a global forest age map. The base age
372 map was derived from biomass through age-biomass curve similarly as conducted in tropical
373 regions in ref.[49] This age map does not cover the northern region beyond 35 N. We filled the
374 missing northern region with a North American age map [50] and a second age map covers
375 China[51]. Remaining missing pixels were further filled with the age map derived from MODIS
376 disturbance observations. For the final step, we filled the remaining pixels with the GFAD V1.1

age map[49]. GFAD V1.1 has 15 age classes and 4 plant functional types (PFTs). We choose the middle value of each age class and estimated the age as the average among different PFTs.

Detailed information of all ancillary variables is listed in Supplementary Table 1. To stay coherent, we re-gridded each map to a common 1 km x 1 km grid through the nearest neighbourhood method.

Building predicting models

We investigated the performance of the allometric scaling and three non-parametric models: RF, ANN and MARS. Allometric upscaling relates root biomass to shoot biomass in the form of $R \propto S^{\beta}$. RF is an ensemble machine learning method that builds a number of decision trees through training samples[52]. A decision tree is a flow-chart-like structure, where each internal (non-leaf) node denotes a binary test on a predicting variable, each branch represents the outcome of a test, and each leaf (or terminal) node holds a predicted target variable. With a combination of learning trees (models), RF generally increases the overall predicting performance and reduces over-fitting. ANN computes through an interconnected group of nodes, inspired by a simplification of neurons in a brain. MARS is a non-parametric regression method that builds multiple linear regression models across a range of predictors.

Tree shoot biomass from the *in-situ* observation data spans a wider range than shoot biomass per plant derived from global maps ($1x10^{-7}$ to 8800 vs. $7.9x10^{-5}$ to 933 kg/plant). To reduce potential mapping errors, we selected training samples with shoot biomass between $5x10^{-5}$ and 1000 kg/plant. The medians and means of shoot biomass, root biomass and *R:S* from the selected training samples are similar as that from the entire database. Also, to reduce the potential impact of outliers, we analyzed samples with *R:S* falling between the 1st and 99th percentiles, which consists of 9589 samples with *R:S* ranging from 0.05 to 2.47 and a mean of 0.47 and a median of 0.36. Sample filtering slightly deteriorated model performance and had minor impact on the final global root biomass prediction (145 from whole samples vs.142 Pg from filtered data). We chose root biomass as our target variable instead of *R:S* because big and small trees contribute equally to *R:S* while big trees are relatively more important in biomass quantification. In our observation database, we have more samples being small woody plants. A predicting model with an overall good performance will not guarantee a good prediction on woody plants with higher biomass. We, furthermore split the *in-situ* measured shoot biomass into three groups, namely with shoot biomass smaller than 0.1, between 0.1 and 10, and larger than

10 kg/plant. The rationale behind this splitting is: (1), the distribution of *in-situ* measured woody shoot biomass (Figure 2); (2), empirical evidence showing the shift of root shoot allometry with tree size[44 20] ; (3), a better performance on independent validation samples through numerous combinations of splitting trials; (4), tests through weighting samples or resampling samples (e.g., over-sampling using Synthetic Minority Over-sampling Technique) gave no better performance.

Model performances were assessed by 4-fold cross-validation using two criteria: the mean absolute error (MAE), the R-squared value (R^2). MAE quantifies the overall error while R^2 estimates the proportion of variance in root biomass that is captured by the predicting model. We favored the model with a smallest MAE, a highest R^2 and with minimum number of predictors. For non-parametric models, starting from a model with all 47 predictors, we sequentially excluded predictors that did not improve model performance one after another. The order of predictor removing is random. After a combination of trials, the best model is from RF and the final set of predictors include shoot biomass, height, soil nitrogen, pH, bulk density, clay content, sand content, base saturation, cation exchange capacity, vapor pressure, mean annual precipitation, mean annual temperature, aridity and water table depth.

Generation of the global root biomass map

We assumed shoot size and other selected predictors to be drivers of root biomass. Building upon a large set of samples with each field measurement being an outcome of complex local interactions (including within-vegetation competition), we implicitly accounted for sub-pixel variability (e.g., resource competition and responses to environmental conditions) on allometry. Biome class and species were excluded from the pool of predicting variables because they did not improve model performance. We combined the RF model with global maps of selected predicting variables to produce the root biomass map which has a unit of weight per tree. This map was multiplied by tree density at 1-km resolution to obtain the final root biomass map with a unit of weight per area (Supplementary Figure 1).

Uncertainty quantification

We estimated the overall uncertainty of the root biomass estimates through quantifying relative errors caused by predictors at the 1-km resolution, predicting errors associated with RF given correct predicting variables, and errors from upscaling root biomass per tree to root biomass per unit area.

Predictor errors (η_{pred}): We collected 8 additional global predictor datasets (3 shoot biomass, 2

soil and 3 climate datasets) (Supplementary Table 2). We carried out 8 sets of additional predictions replacing the predictors by each of these additional data maps and calculated the standard deviation among 8 predictions for each pixel. The overall predictor errors were expressed in a relative term, that is, the ratio between the standard deviation and the standard prediction (with the GlobBiomass-AGB and other predictors listed in Supplementary Table 1) for each pixel.

RF errors (η_{RF}): The performance of machine learning models is frequently verified through the independent test samples. We carried out 4-fold cross-validation. The RF error is quantified as the relative error (the standard deviation divided by the mean) from 4-fold predictions.

Upscaling errors (η_{up}): Upscaling the root biomass from per tree to per area relies on the tree density map. The upscaling error is set as the relative uncertainty of tree density[25].

At last we propagated these relative errors across the entire root biomass quantification processes assuming these three errors were random and independent. So the errors were assumed to be uncorrelated and the covariation were assumed to be 0. The overall relative errors at the pixel level was calculated through,

$$\eta_{root} = \sqrt{\eta_{pred}^2 + \eta_{RF}^2 + \eta_{up}^2} \qquad (1)$$

Uncertainty at the global or biome scale (σ_{biome}) is quantified through expanding calculating area and propagating the relative errors at the pixel level,

$$\sigma_{biome} = \sqrt{\sum_{i=1}^{N} (BR_i \eta_{root_i})^2} \qquad (2)$$

where BR is the total root biomass (in unit of weight) in each forested pixel and N is the number of pixels within biome boundaries (or all forested pixels when calculate the global total). η_{root_i} is the relative uncertainty in quantifying root biomass for the *ith* pixel.

Relative importance of predicting variables

The impact of predictors on predicting *R:S* was estimated through the Spearman's rank-order correlation at both the global and biome scales. We log-transformed the *R:S* and shoot biomass before standardizing these datasets. Partial dependence plot[53] tells the marginal effect of

one predictor have on root biomass from a machine learning model, and serves as a supplement to the Spearman correlation.

Reference

1 Jackson, R. B., Mooney, H. A. & Schulze, E. D. A global budget for fine root biomass, surface area, and nutrient contents. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 7362-7366, doi:10.1073/pnas.94.14.7362 (1997).

2 Saugier, B., Roy, J. & Mooney, H. A. Estimations of global terrestrial productivity: converging toward a single number? In: Terrestrial Global Productivity (eds Roy J, Saugier B, Mooney HA), pp. 543–556. Academic Press, San Diego. (2001).

3 Robinson, D. Implications of a large global root biomass for carbon sink estimates and for soil carbon dynamics. *Proceedings of the Royal Society B-Biological Sciences* **274**, 2753-2759, doi:10.1098/rspb.2007.1012 (2007).

4 Warren, J. M. *et al.* Root structural and functional dynamics in terrestrial biosphere models - evaluation and recommendations. *New Phytologist* **205**, 59-78, doi:10.1111/nph.13034 (2015).

5 Wang, Y. Y., Xie, Z. H. & Jia, B. H. Incorporation of a dynamic root distribution into CLM4.5: Evaluation of carbon and water fluxes over the Amazon. *Advances in Atmospheric Sciences* **33**, 1047-1060, doi:10.1007/s00376-016-5226-8 (2016).

6 Santoro, M. e. a. GlobBiomass - global datasets of forest biomass. PANGAEA, https://doi.org/10.1594/PANGAEA.894711. (2018).

7 GlobalForestWatch. (https://data.globalforestwatch.org/datasets/8f93a6f94a414f9588ce4657a39c59ff_1, 2019).

8 Taylor, B. N., Beidler, K. V., Cooper, E. R., Strand, A. E. & Pritchard, S. G. Sampling volume in root studies: the pitfalls of under-sampling exposed using accumulation curves. *Ecology Letters* **16**, 862-869, doi:10.1111/ele.12119 (2013).

9 Jackson, R. B. *et al.* A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**, 389-411, doi:10.1007/bf00333714 (1996).

10 Enquist, B. J. & Niklas, K. J. Global allocation rules for patterns of biomass partitioning in seed plants. *Science* **295**, 1517-1520, doi:10.1126/science.1066360 (2002).

11 McCarthy, M. C. & Enquist, B. J. Consistency between an allometric approach and optimal partitioning theory in global patterns of plant biomass allocation. *Functional Ecology* **21**, 713-720, doi:10.1111/j.1365-2435.2007.01276.x (2007).

12 Robinson, D. Scaling the depths: below-ground allocation in plants, forests and biomes. *Functional Ecology* **18**, 290-295, doi:10.1111/j.0269-8463.2004.00849.x (2004).

13 Cairns, M. A., Brown, S., Helmer, E. H. & Baumgardner, G. A. Root biomass allocation in the world's upland forests. *Oecologia* **111**, 1-11, doi:10.1007/s004420050201 (1997).

14 Jiang, Y. T. & Wang, L. M. Pattern and control of biomass allocation across global forest ecosystems. *Ecology and Evolution* **7**, 5493-5501, doi:10.1002/ece3.3089 (2017).

15 Niklas, K. J. Modelling below- and above-ground biomass for non-woody and woody plants. *Annals of Botany* **95**, 315-321, doi:10.1093/aob/mci028 (2005).
16 Yunjian Luo *et al.* ChinAllomeTree 1.0. (2018).
17 Cheng, D. L. & Niklas, K. J. Above- and below-ground biomass relationships across 1534 forested communities. *Annals of Botany* **99**, 95-102, doi:10.1093/aob/mcl206 (2007).
18 Luo, Y. J., Wang, X. K., Zhang, X. Q., Booth, T. H. & Lu, F. Root:shoot ratios across China's forests: Forest type and climatic effects. *Forest Ecology and Management* **269**, 19-25, doi:10.1016/j.foreco.2012.01.005 (2012).
19 Moser, G. *et al.* Elevation effects on the carbon budget of tropical mountain forests (S Ecuador): the role of the belowground compartment. *Global Change Biology* **17**, 2211-2226, doi:10.1111/j.1365-2486.2010.02367.x (2011).
20 Ledo, A. *et al.* Tree size and climatic water deficit control root to shoot ratio in individual trees globally. *New Phytologist* **217**, 8-11, doi:10.1111/nph.14863 (2018).
21 Reich, P. B. *et al.* Temperature drives global patterns in forest biomass distribution in leaves, stems, and roots. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13721-13726, doi:10.1073/pnas.1216053111 (2014).
22 Olson, D. M. & Dinerstein, E. The Global 200: Priority ecoregions for global conservation. (PDF file) Annals of the Missouri Botanical Garden 89:125-126. -The Nature Conservancy, USDA Forest Service and U.S. Geological Survey, based on Bailey, Robert G. 1995. Description of the ecoregions of the United States (2nd ed.). Misc. Pub. No. 1391, Map scale 1:7,500,000. USDA Forest Service. 108pp. -The Nature Conservancy (2003), based on Wiken, E.B. (compiler). 1986. Terrestrial ecozones of Canada. Ecological Land Classification Series No. 19. Environment Canada, Hull, Que. 26 pp. + map. (2002).
23 Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850-853, doi:10.1126/science.1244693 (2013).
24 Santoro, M., Cartus, O., Mermoz, S., Bouvet, A., Le Toan, T., Carvalhais, N., Rozendaal, D., Herold, M., Avitabile, V., Quegan, S., Carreiras, J., Rauste, Y., Balzter, H., Schmullius, C. and Seifert, F. M.. . A detailed portrait of the forest aboveground biomass pool for the year 2010 obtained from multiple remote sensing observations. Geophysical Research Abstracts, 20, EGU2018-18932, EGU General Assembly 2018. (2018).
25 Crowther, T. W. *et al.* Mapping tree density at a global scale. *Nature* **525**, 201-+, doi:10.1038/nature14967 (2015).
26 Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *Journal of Geophysical Research-Biogeosciences* **116**, doi:10.1029/2011jg001708 (2011).
27 Shangguan, W., Dai, Y. J., Duan, Q. Y., Liu, B. Y. & Yuan, H. A global soil data set for earth system modeling. *Journal of Advances in Modeling Earth Systems* **6**, 249-263, doi:10.1002/2013ms000293 (2014).
28 Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302-4315, doi:10.1002/joc.5086 (2017).
29 Trabucco, A. & Zomer, R. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2. figshare. Dataset. (2019).

30 Fan, Y., Li, H. & Miguez-Macho, G. Global Patterns of Groundwater Table Depth. *Science* **339**, 940-943, doi:10.1126/science.1229881 (2013).
31 Staver, A. C., Archibald, S. & Levin, S. A. The Global Extent and Determinants of Savanna and Forest as Alternative Biome States. *Science* **334**, 230-232, doi:10.1126/science.1210465 (2011).
32 Cheng Meng *et al.* Global soil acidification impacts on belowground processes. Environmental Research Letters, Volume 14, Number 7. (2019).
33 Schepaschenko, D. *et al.* Improved Estimates of Biomass Expansion Factors for Russian Forests. *Forests* **9**, doi:10.3390/f9060312 (2018).
34 Waring, B. G. & Powers, J. S. Overlooking what is underground: Root:shoot ratios and coarse root allometric equations for tropical forests. *Forest Ecology and Management* **385**, 10-15, doi:10.1016/j.foreco.2016.11.007 (2017).
35 Mokany, K., Raison, R. J. & Prokushkin, A. S. Critical analysis of root: shoot ratios in terrestrial biomes. *Global Change Biology* **12**, 84-96, doi:10.1111/j.1365-2486.2005.001043.x (2006).
36 Litton, C. M., Raich, J. W. & Ryan, M. G. Carbon allocation in forest ecosystems. *Global Change Biology* **13**, 2089-2109, doi:10.1111/j.1365-2486.2007.01420.x (2007).
37 West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122-126, doi:10.1126/science.276.5309.122 (1997).
38 West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure and allometry of plant vascular systems. *Nature* **400**, 664-667 (1999).
39 Fan, Y., Miguez-Macho, G., Jobbagy, E. G., Jackson, R. B. & Otero-Casal, C. Hydrologic regulation of plant rooting depth. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 10572-10577, doi:10.1073/pnas.1712381114 (2017).
40 Friend, A. D. *et al.* Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO2. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3280-3285, doi:10.1073/pnas.1222477110 (2014).
41 Bloom, A. A., Exbrayat, J. F., van der Velde, I. R., Feng, L. & Williams, M. The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 1285-1290, doi:10.1073/pnas.1515160113 (2016).
42 Falster, D. S. *et al.* BAAD: a Biomass And Allometry Database for woody plants. *Ecology* **96**, 1445-1445, doi:10.1890/14-1889.1 (2015).
43 Schepaschenko, D. *et al.* A dataset of forest biomass structure for Eurasia. *Scientific Data* **4**, doi:10.1038/sdata.2017.70 (2017).
44 Poorter, H. *et al.* How does biomass distribution change with size and differ among species? An analysis for 1200 plant species from five continents. *New Phytologist* **208**, 736-749, doi:10.1111/nph.13571 (2015).
45 Iversen, C. M. *et al.* A global Fine-Root Ecology Database to address below-ground challenges in plant ecology. *New Phytologist* **215**, 15-26, doi:10.1111/nph.14486 (2017).

46 Santoro, M. *et al.* Forest growing stock volume of the northern hemisphere: Spatially explicit estimates for 2010 derived from Envisat ASAR. *Remote Sensing of Environment* **168**, 316-334, doi:10.1016/j.rse.2015.07.005 (2015).

47 Tuanmu, M. N. & Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography* **23**, 1031-1045, doi:10.1111/geb.12182 (2014).

48 Hagolle, O. *et al.* Quality assessment and improvement of temporally composited products of remotely sensed imagery by combination of VEGETATION 1 and 2 images. *Remote Sensing of Environment* **94**, 172-186, doi:10.1016/j.rse.2004.09.008 (2005).

49 Poulter, B. A., Luiz; Andela, Niels; Bellassen, Valentin; Ciais, Philippe; Kato, Tomomichi; Lin, Xin; Nachin, Baatarbileg; Luyssaert, Sebastiaan; Pederson, Niel; Peylin, Philippe; Piao, Shilong; Pugh, Tom; Saatchi, Sassan; Schepaschenko, Dmitry; Schelhaas, Martjan; Shivdenko, Anatoly The global forest age dataset and its uncertainties (GFADv1.1). NASA National Aeronautics and Space Administration, PANGAEA, https://doi.org/10.1594/PANGAEA.897392. (2019).

50 Pan, Y. *et al.* Age structure and disturbance legacy of North American forests. *Biogeosciences* **8**, 715-732, doi:10.5194/bg-8-715-2011 (2011).

51 Zhang, Y., Yao, Y. T., Wang, X. H., Liu, Y. W. & Piao, S. L. Mapping spatial distribution of forest age in China. *Earth and Space Science* **4**, 108-116, doi:10.1002/2016ea000177 (2017).

52 Breiman, L. Random forests. *Machine Learning* **45**, 5-32, doi:10.1023/a:1010933404324 (2001).

53 Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition, Section 10.13.2, Springer*. (2009).

Acknowledgements

Y.H., D.S.G and P.C. received support from the European Research Council Synergy project SyG-2013-610028 IMBALANCE-P and P.C. and Y.H. from the ANR CLAND Convergence Institute. H.Y., D.S. and M. S. were funded through the ESA Climate Change Initiative BIOMASS project. Collecting Russian data were supported by The Russian Science Foundation (project no. 19-77-30015).

Author contributions

Y.H. and P.C. designed this study. Y. H., P.C., M.S., J.C and D.S. collected the data. D.M., P.C., M.S., J.C., Y.C. and Y.H. discussed analyzing methods. Y.H. conducted the analysis and drafted the manuscript. All authors discussed the results and contributed to the manuscript.

Code availability

Calculations were conducted through Python 2.7.15 and ferret 6.72. The code is available upon request.

Data availability

The datasets generated in this study are available from the corresponding author on reasonable request.